

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
2 October 2003 (02.10.2003)

PCT

(10) International Publication Number  
**WO 03/080795 A2**

(51) International Patent Classification<sup>7</sup>: C12N

(21) International Application Number: PCT/US02/25485

(22) International Filing Date: 9 August 2002 (09.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/311,261 9 August 2001 (09.08.2001) US

(71) Applicant (for all designated States except US): HYSEQ, INC. [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94085 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). YANG, Yonghong [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). WANG, Zhiwei [CN/US]; 836 Alturas Ave #B36, Sunnyvale, CA 94085 (US). WENG, Gezhi [US/US]; 415 Moraga Avenue, Piedmont, CA 94611 (US). MA, Yunqing [CN/US]; 280 W. California Avenue #206, Sunnyvale, CA 94086 (US).

(74) Agents: RIN-LAURES, Li-Hsien et al.; Hyseq, Inc., 670 Almanor Avenue, Sunnyvale, CA 94085 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

*without international search report and to be republished upon receipt of that report with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: NOVEL NUCLEIC ACIDS AND SECRETED POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.



WO 03/080795 A2

## NOVEL NUCLEIC ACIDS AND SECRETED POLYPEPTIDES

### 1. CROSS REFERENCE TO RELATED APPLICATIONS

5           This application is a continuation-in-part application of U.S. Application Serial No. 09/552,317 filed April 25, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/488,725 filed January 21, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784; U.S. Application Serial No. 09/491,404  
10       filed January 25, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 785; U.S. Application Serial No. 09/560,875 filed April 27, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/496,914 filed February 03, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787;  
15       U.S. Application Serial No. 09/577,409 filed May 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/515,126 filed February 28, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788; U.S. Application Serial No. 09/574,454 filed May 19, 2000 entitled "Novel Contigs  
20       Obtained from Various Libraries", Attorney Docket No. 789CIP which in turn is a continuation-in-part application of U.S. Application Serial No. 09/519,705 filed March 07, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 789; U.S. Application Serial No. 09/649,167 filed August 23, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790CIP, which in turn is a  
25       continuation-in-part application of U.S. Application Serial No. 09/540,217 filed March 31, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790; U.S. Application Serial No. 09/770,160 filed January 26, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791CIP, which is in turn a continuation-in-part application of U.S. Application Serial No. 09/552,929 filed April 18,  
30       2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791; and U.S. Application Serial No. 09/577,408 filed May 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 792; all of which are incorporated herein by reference in their entirety.



## 2. BACKGROUND OF THE INVENTION

### 2.1 TECHNICAL FIELD

5       The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

### 2.2 BACKGROUND

10       Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, circulating soluble factors, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence  
15 of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making  
20 available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in,  
25 for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

## 3. SUMMARY OF THE INVENTION

30       The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize

one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1-1041, or 2083-2534 and are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases or unknown. In the amino acids provided in the Sequence Listing, \* corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1-1041, or 2083-2534 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1-1041, or 2083-2534. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1-1041, or 2083-2534 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-1041, or 2083-2534. The sequence information can be a segment of any one of SEQ ID NO: 1-1041, or 2083-2534 that uniquely identifies or represents the sequence information of SEQ ID NO: 1-1041, or 2083-2534.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information are provided on a nucleic acid array to detect the polynucleotide that contains the segment. The

array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-1041, or 2083-2534 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-1041, or 2083-2534 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., *Science* 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 1-1041, or 2083-2534; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1-1041, or 2083-2534; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO: 1-1041, or 2083-2534. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1-1041, or 2083-2534; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in SEQ ID NO: 1-1041, or 2083-2534; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in SEQ ID NO: 1042-2082, or 2535-2986, or Tables 3, 5, 6, or 8.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having  
5 a nucleotide sequence set forth in SEQ ID NO: 1-1041, or 2083-2534; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological  
10 activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such  
15 as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium  
20 under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such processes is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques  
25 include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence  
30 of the particular cell or tissue mRNA in a sample using, e.g., *in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and

exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions.

The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein.

5 Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (e.g., bind to) the polypeptides of the invention. The invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives  
10 expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound that binds to a polypeptide of the invention is identified.

The methods of the invention also provide methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals  
15 exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can affect such modulation either on the level of target gene/protein expression or target protein activity.

20 The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Table 2); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Table 4). If no homology is set forth for a sequence, then the polypeptides  
25 and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

#### 4. DETAILED DESCRIPTION OF THE INVENTION

##### 30 4.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule.

5 Likewise "immunologically active" or "immunological activity" refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of  
10 secretory or enzymatic molecules as part of a normal or disease process.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only certain portion(s) of the nucleic acids bind or it  
15 may be "complete" such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ  
20 line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source  
25 from which GSCs and ES cells are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides  
30 which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences

(inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonucleotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and N is A, C, G, or T (U) or unknown. It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NO: 1-1041, or 2083-2534.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal



DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-1041, or 2083-2534. The sequence information can be a segment of any one of SEQ ID NO: 1-1041, or 2083-2534 that uniquely identifies or represents the sequence information of that sequence of SEQ ID NO: 1-1041, or 2083-2534, or those segments identified in Tables 3, 5, 6, and 8. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because  $4^{20}$  possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ( $1 \div 4^{25}$ ) times the increased probability for mismatch at each nucleotide position ( $3 \times 25$ ). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence.

While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

5 The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a  
10 stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any  
15 polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

20 The term "translated protein coding portion" means a sequence which encodes for the full-length protein which may include any leader sequence or any processing sequence.

The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide  
25 may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

30 The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or

substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant"(or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions,-deletions, and substitutions, created using, *e g.*, recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such

alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate

5 polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological

10 macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

15 The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (*e.g.*, nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not

20 encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (*e.g.*, yeast) expression systems. As a product, "recombinant

25 microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

30 The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or

enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell.

- 5 Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

- 10 The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or
- 15 elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

- 20 The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are
- 25 also intended to include proteins containing non-typical signal sequences (*e.g.* Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2): 134 -143) and factors released from damaged cells (*e.g.* Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

- 30 Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M NaHPO<sub>4</sub>, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligonucleotides), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" or "substantially similar" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (i.e., the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, e.g., mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more than 5% (95% sequence identity). Substantially equivalent, e.g., mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. Substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, the nucleotide sequence has at least about 65% identity, more preferably at least

about 75% identity, more preferably at least about 80% sequence identity, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least about 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. For the purposes of the present invention,

- 5 sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a new stop codon) should be disregarded. Sequence identity may be determined, *e.g.*, using the Jotun Hein method (Hein, J. (1990) *Methods Enzymol.* 183:626-645).
- 10 Identity between sequences can also be determined by other methods known in the art, *e.g.* by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

- The term "transformation" means introducing DNA into a suitable host cell so that
- 15 the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

- 20 As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid
- 25 molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

30

#### 4.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1-1041, or 2083-2534; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO: 1-1041, or 2083-2534; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1-1041, or 2083-2534. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotide sequences of SEQ ID NO: 1-1041, or 2083-2534; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing, or Table 8; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 1042-2082, or 2535-2986 (for example, as set forth in Tables 3, 5, 6, or 8). Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include entire coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1-1041, or 2083-2534 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1-1041, or 2083-2534 or a portion thereof as a probe. Alternatively, the polynucleotides of



SEQ ID NO: 1-1041, or 2083-2534 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above.

Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, more typically at least about 85%, 86%, 87%, 88%, 89%, more typically at least about 90%, 91%, 92%, 93%, 94%, and even more typically at least about 95%, 96%, 97%, 98%, 99% sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO: 1-1041, or 2083-2534, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, *e.g.* 15, 17, or 20 nucleotides or more that are selective for (*i.e.* specifically hybridize to) any one of the polynucleotides of the invention are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1-1041, or 2083-2534, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NO: 1-1041, or 2083-2534 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology results for the nucleic acids of the present invention, including SEQ ID NO: 1-1041, or 2083-2534 can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST (Basic Local Alignment Search Tool) program is used to search for local sequence alignments (Altschul, S.F. J Mol. Evol. 36 290-300 (1993) and  
5 Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using FASTXY algorithm may be performed.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a  
10 suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

15 The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic  
20 acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative  
25 choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal  
30 fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for

intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., *supra*, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention could be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such

polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature  
5 protein coding sequences corresponding to any one of SEQ ID NO: 1-1041, or 2083-2534, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other  
10 nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a  
15 polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or  
20 eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-1041, or 2083-2534 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral  
25 vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-1041, or 2083-2534 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present  
30 invention. The following vectors are provided by way of example: Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene), pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic:

pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al.,

5 *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly.

Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are

10 situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two  
15 appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of  
20 replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among  
25 others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, e.g., stabilization or  
30 simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic

selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., Nat. Biotech 17, 870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intra-muscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

#### 4.3 ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1-1041, or 2083-2534, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a

sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of SEQ ID NO: 1-1041, or 2083-2534 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO: 1-1041, or 2083-2534 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO: 1-1041, or 2083-2534, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of an mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of an mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine,

1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a protein according to the invention to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\alpha$ -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15:



6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

#### 5           4.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in  
10 Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of an mRNA. A ribozyme having specificity for a nucleic acid of the invention can be designed based upon the nucleotide sequence of a DNA disclosed herein (*i.e.*, SEQ ID NO: 1-1041, or 2083-2534). For example, a derivative of Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the  
15 active site is complementary to the nucleotide sequence to be cleaved in a mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, mRNA of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

20           Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

25           In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup  
30 *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The

synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of the invention can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of the invention can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.*

86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

#### 4.5 HOSTS

10 The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are  
15 in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (*e.g.*, by homologous recombination) to provide increased polypeptide expression by replacing, in  
20 whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also  
25 contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (*e.g.*, *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification  
30 of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by

calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell* 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used,

as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

5 Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli*, *Bacillus subtilis*, *Salmonella typhimurium*, or any bacterial  
10 strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

15 In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory  
20 sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, and regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein  
25 produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

30 The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory

element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

#### 4.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 1042-2082, or 2535-2986 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NO: 1-1041, or 2083-2534 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID NO: 1-1041, or 2083-2534 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 1042-2082, or 2535-2986 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as

SEQ ID NO: 1042-2082, or 2535-2986 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, 86%, 87%, 88%, 89%, at least about 90%, 91%, 92%, 93%, 94%, typically at least about 95%, 96%, 97%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 1042-2082, or 2535-2986.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., *Bio/Technology* 10, 773-778 (1992) and in R. S. McDowell, et al., *J. Amer. Chem. Soc.* 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites. Fragments are also identified in Tables 3, 5, 6, and 8.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The predicted signal sequence is set forth in Table 6. The mature form of such protein may be obtained and confirmed by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell and sequencing of the cleaved product. One of skill in the art will recognize that the actual cleavage site may be different than that predicted in Table 6. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide

fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

5           A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological  
10   properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the  
15   development of antibodies.

          The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally  
20   does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

          The invention also relates to methods for producing a polypeptide comprising  
25   growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The  
30   polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.



In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, e.g., Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*. Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 1042-2082, or 2535-2986.

The protein of the invention may also be expressed as a product of transgenic animals, e.g., as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to

alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, *e.g.*, U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of

maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, e.g., silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

#### 4.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE

##### IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP

(Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), Pfam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobicity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), incorporated herein by reference).

10 polypeptide sequences were examined by a proprietary algorithm, SeqLoc that separates the proteins into three sets of locales: intracellular, membrane, or secreted. This prediction is based upon three characteristics of each polypeptide, including percentage of cysteine residues, Kyte-Doolittle scores for the first 20 amino acids of each protein, and Kyte-Doolittle scores to calculate the longest hydrophobic stretch of the said protein. Values of

15 predicted proteins are compared against the values from a set of 592 proteins of known cellular localization from the Swissprot database (<http://www.expasy.ch/sprot>). Predictions are based upon the maximum likelihood estimation.

The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

#### 4.7 CHIMERIC AND FUSION PROTEINS

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to

25 another polypeptide. Within a fusion protein the polypeptide according to the invention can correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the invention. In another embodiment, a fusion protein comprises at least two biologically active portions of a protein according to the invention. Within the fusion protein, the term

30 "operatively linked" is intended to indicate that the polypeptide according to the invention and the other polypeptide are fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus, or to the middle.

For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein.

In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e.,  
5 glutathione S-transferase) sequences.

In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprise one or more domains fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical  
10 compositions and administered to a subject to inhibit an interaction between a ligand and a protein of the invention on the surface of a cell, to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as  
15 modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies in a subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

A chimeric or fusion protein of the invention can be produced by standard  
20 recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic  
25 ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.)  
30 CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be

cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

#### 4.8 GENE THERAPY

5 Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more  
10 particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992).  
15 Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then  
20 be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

25 Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

30 The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of

the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, e.g., inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple

deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

#### 4.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model



systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The

homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

#### 4.10 USES AND BIOLOGICAL ACTIVITY

5       The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of  
10   DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate  
15   variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding  
20   proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

25       The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

##### 4.10.1 RESEARCH USES AND UTILITIES

30       The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on

gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

#### 4.10.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or

amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

#### 4.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of

mouse and human interleukin- $\gamma$ , Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine

- 5 Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in
- 10 Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C.
- 15 and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in:

- 20 Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411,
- 25 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

#### 4.10.4 STEM CELL GROWTH FACTOR ACTIVITY

- 30 A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotent or

pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells.

Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotent/pluripotent stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotent/pluripotent mRNA to create

cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., *Differentiation*, 48: 173-182, (1991); Klug et al., *J. Clin. Invest.*, 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering* eds. Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

*In vitro* cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. *Proc. Natl. Acad. Sci. U.S.A.*, 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the

invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

#### 4.10.5 HEMATOPOIESIS REGULATING ACTIVITY

5 A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, 10 thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in 15 supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and 20 therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or 25 heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

30 Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.



- Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994;
- 5 Hiramama et al., *Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., *Experimental Hematology* 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In *Culture of Hematopoietic Cells*.
- 10 R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc.,
- 15 New York, N.Y. 1994.

#### 4.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and

20 tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have

25 prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming

30 cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast

activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from

chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

#### **4.10.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY**

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and

disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, *Leishmania* spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also be useful in the treatment of allergic reactions and conditions (e.g., anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., *Toxicology* 125: 59-66, 1998), skin prick test (Hoffmann et al., *Allergy* 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., *Arch. Toxicol.* 73: 501-9), and murine local lymph node assay (Kimber et al., *J. Toxicol. Environ. Health* 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of

an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing  
5 non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without  
10 limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition  
15 as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may  
20 avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in  
25 humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., *Science* 257:789-792 (1992) and Turka et al., *Proc. Natl. Acad. Sci USA*, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed.,  
30 *Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self-tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosus in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and  $\beta_2$  microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro

antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

#### 4.10.8 ACTIVIN/INHIBIN ACTIVITY



A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., *Endocrinology* 91:562-572, 1972; Ling et al., *Nature* 321:779-782, 1986; Vale et al., *Nature* 321:776-779, 1986; Mason et al., *Nature* 318:659-663, 1985; Forage et al., *Proc. Natl. Acad. Sci. USA* 83:3091-3095, 1986.

#### 4.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to

tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: *Current Protocols in Immunology*, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. *J. Clin. Invest.* 95:1370-1376, 1995; Lind et al. *APMIS* 103:140-146, 1995; Muller et al. *Eur. J. Immunol.* 25:1744-1748; Gruber et al. *J. of Immunol.* 152:5860-5867, 1994; Johnston et al. *J. of Immunol.* 153:1762-1768, 1994.

#### 4.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostasis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., *J. Clin. Pharmacol.* 26:131-140, 1986; Burdick et al., *Thrombosis*

Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

#### 4.10.11 CANCER DIAGNOSIS AND THERAPY

5 Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing  
10 malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation,  
15 inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies  
20 including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal  
25 neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central  
30 nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Kaposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention

(including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

*In vitro* models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987)

Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

#### 4.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14. Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

#### 4.10.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product  
5 libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science* 282:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide  
10 and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol.* 9(3):205-23 (1998); Hruby  
15 et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then  
20 tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The  
25 toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

#### 4.10.14 ASSAY FOR RECEPTOR ACTIVITY

30 The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening

assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.*, increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The responses of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications *i.e.* phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

#### 4.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an



inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis,

- 5 complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis,
- 10 acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflammation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic myelogenous leukemia or in the prevention of premature labor secondary to
- 15 intrauterine infections.

#### 4.10.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of

20 the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

25

#### 4.10.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of

30 therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include

but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or

5 compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or  
10 injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration  
15 associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency,  
20 Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

25 (vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various  
30 etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival

or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

- (i) increased survival time of neurons in culture;
- (ii) increased sprouting of neurons in culture or *in vivo*;
- 5 (iii) increased production of a neuron-associated molecule in culture or *in vivo*,  
e.g., choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or
- (iv) decreased symptoms of neuron dysfunction *in vivo*.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method  
10 set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor  
15 neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, e.g., weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor  
20 neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory  
25 Neuropathy (Charcot-Marie-Tooth Disease).

#### 4.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing,  
30 infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution,

change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s);  
5 effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of  
10 the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

#### 4.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential  
20 predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in  
25 humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate  
30 fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that

hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

#### 4.10.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et al., 1983, *Science*, 219:56, or by B. Waksman et al., 1963, *Int. Arch. Allergy Appl. Immunol.*, 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed *Mycobacterium tuberculosis* in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed *Mycobacterium tuberculosis* in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of *Mycobacterium* CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

#### 4.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

5

#### 4.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01  $\mu\text{g/kg}$  to 100  $\text{mg/kg}$  of body weight, with the preferred dose being about 0.1  $\mu\text{g/kg}$  to 10  $\text{mg/kg}$  of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

#### 4.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other

materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration.

The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF $\alpha$ , TNF $\beta$ , TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF- $\alpha$  and TGF- $\beta$ ), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound

sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

#### 4.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in



fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

#### 4.12.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical

composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from  
5 about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally  
10 acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride  
15 Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or  
20 physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such  
25 carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable  
30 excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired,

disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such

as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics.

Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, *e.g.* polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable

matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides,

diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like.

Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

- 5           The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient.
- 10   Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to
- 15   practice the method of the present invention should contain about 0.01  $\mu$ g to about 100 mg (preferably about 0.1  $\mu$ g to about 10 mg, more preferably about 0.1  $\mu$ g to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition
- 20   topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically
- 25   useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active
- 30   ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above-mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet

derived growth factor (PDGF), transforming growth factors (TGF- $\alpha$  and TGF- $\beta$ ), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either *in vivo* or *ex vivo* into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes.

#### 4.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be



estimated initially from appropriate *in vitro* assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the  $IC_{50}$  as  
5 determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic  
10 efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the  $LD_{50}$  (the dose lethal to 50% of the population) and the  $ED_{50}$  (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between  $LD_{50}$  and  $ED_{50}$ . Compounds which exhibit high therapeutic  
15 indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the  $ED_{50}$  with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of  
20 administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in*  
25 *vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of  
30 the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 µg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 µg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

#### 4.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

#### 4.13 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F<sub>ab</sub>, F<sub>ab</sub>' and F<sub>(ab)2</sub> fragments, and an F<sub>ab</sub> expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG<sub>1</sub>, IgG<sub>2</sub>, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for

polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in  
5 SEQ ID NO: 1042-2082, or 2535-2986, or Tables 3, 5, 6, or 8, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred  
10 epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a surface region of the protein, e.g., a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of  
15 a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, e.g., Hopp and  
20 Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog  
25 thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (i.e., able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence  
30 identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine

binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), *Antibodies A Laboratory Manual*; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the

invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full-length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein.

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known

in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

#### 4.13.1 POLYCLONAL ANTIBODIES

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface-active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific

antigen which is the target of the immune globulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

#### 4.13.2 MONOCLONAL ANTIBODIES

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen-binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256, 495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas

typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia.

Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, *J. Immunol.*, 133:3001 (1984);

Brodeur et al., *Monoclonal Antibody Production Techniques and Applications*, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.*, 107, 220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as

a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA  
5 also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted  
10 for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

#### 4.13.3 HUMANIZED ANTIBODIES

15 The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab',  
20 F(ab')<sub>2</sub> or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321, 522-525 (1986); Riechmann et al., Nature, 332, 323-327 (1988); Verhoeyen et al., Science, 239, 1534-1536 (1988)), by substituting  
25 rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539). In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise  
30 substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion



of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, *Curr. Op. Struct. Biol.*, 2, 593-596 (1992)).

#### 5           4.13.4 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human  
10 B-cell hybridoma technique (see Kozbor, et al., 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. *Proc Natl Acad Sci USA* 80,  
15 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227, 381 (1991); Marks et al., *J. Mol. Biol.*, 222:581 (1991)). Similarly, human antibodies can be made by  
20 introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806;  
25 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (*Bio/Technology* 10, 779-783 (1992)); Lonberg et al. (*Nature* 368, 856-859 (1994)); Morrison (*Nature* 368, 812-13 (1994)); Fishwild et al, (*Nature Biotechnology* 14, 845-51 (1996)); Neuberger (*Nature Biotechnology* 14, 826 (1996)); and Lonberg and Huszar (*Intern. Rev. Immunol.* 13, 65-93 (1995)).

30 Human antibodies may additionally be produced using transgenic nonhuman animals that are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains

in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then  
5 obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse<sup>TM</sup> as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells that secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after  
10 immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for  
15 example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent  
20 rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

25 A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The  
30 hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that

binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

#### 4.13.5 FAB FRAGMENTS AND SINGLE CHAIN ANTIBODIES

5 According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of  $F_{ab}$  expression libraries (see e.g., Huse, et al., 1989 Science 246, 1275-1281) to allow rapid and effective identification of monoclonal  $F_{ab}$  fragments with the desired specificity for a protein or  
10 derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an  $F_{(ab)2}$  fragment produced by pepsin digestion of an antibody molecule; (ii) an  $F_{ab}$  fragment generated by reducing the disulfide bridges of an  $F_{(ab)2}$  fragment; (iii) an  $F_{ab}$  fragment generated by the treatment of the antibody molecule with papain and a reducing  
15 agent and (iv)  $F_v$  fragments.

#### 4.13.6 BISPECIFIC ANTIBODIES

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of  
20 the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two  
25 immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305, 537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished  
30 by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10, 3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion

preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., *Methods in Enzymology*, 121, 210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers that are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full-length antibodies or antibody fragments (e.g.  $F(ab')_2$  bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., *Science* 229, 81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate  $F(ab')_2$  fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The  $Fab'$  fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the  $Fab'$ -TNB derivatives is then reconverted to the  $Fab'$ -thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other  $Fab'$ -TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally,  $Fab'$  fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175, 217-225 (1992) describe the production of a fully humanized bispecific antibody  $F(ab')_2$  molecule. Each  $Fab'$  fragment was separately secreted from *E. coli* and subjected to directed chemical

coupling *in vitro* to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5), 1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90, 6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain ( $V_H$ ) connected to a light-chain variable domain ( $V_L$ ) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the  $V_H$  and  $V_L$  domains of one fragment are forced to pair with the complementary  $V_L$  and  $V_H$  domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152, 5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., *J. Immunol.* 147, 60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG ( $Fc\gamma R$ ), such as  $Fc\gamma RI$  (CD64),  $Fc\gamma RII$  (CD32) and  $Fc\gamma RIII$  (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

#### 4.13.7 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells  
5 (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include  
10 iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

#### 4.13.8 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector  
15 function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176, 1191-1195 (1992) and Shopes, J. Immunol., 148, 2918-2922 (1992).  
20 Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53, 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al.,  
25 Anti-Cancer Drug Design, 3, 219-230 (1989).

#### 4.13.9 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active  
30 toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used

include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include  $^{212}\text{Bi}$ ,  $^{131}\text{I}$ ,  $^{131}\text{In}$ ,  $^{90}\text{Y}$ , and  $^{186}\text{Re}$ .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

#### 4.14 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the

presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known  
5 methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means  
10 chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database  
15 application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g. text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NO: 1-1041, or 2083-2534 or  
20 a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NO: 1-1041, or 2083-2534 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow  
25 demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein-encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions  
30 and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the



present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means  
5 having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access  
10 manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target  
15 sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available  
20 algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The  
25 most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally  
30 selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include,

but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

#### 4.15 TRIPLE HELIX FORMATION

5 In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple  
10 helix-see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 15241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization  
15 blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

#### 20 4.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

25 In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization  
30 conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

5 In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods  
10 employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science  
15 Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or  
20 membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is  
25 compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies  
30 of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

#### 4.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

#### 4.18 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NO: 1-1041, or 2083-2534, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

- (a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and  
5 detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting  
10 the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression  
15 of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to  
20 activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in  
25 the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

30 For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed"

when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al.,  
5 Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or  
10 EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple  
15 helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix -  
20 see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Okano, J. Neurochem. 56, 560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks  
25 translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention  
30 can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

#### 4.19 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NO: 1-1041, or 2083-2534. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from any of the nucleotide sequences SEQ ID NO: 1-1041, or 2083-2534 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, *in situ* hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well-known genetic and/or chromosomal mapping techniques. These techniques include *in situ* hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent *in situ* hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal

map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

#### 4.20 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6), 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8), 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed CovaLink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridgeheads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) Anal. Biochem. 198(1) 138-42).

The use of CovaLink NH-strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen *et al.*, (1991). In this technology, a phosphoramidate bond is employed (Chu *et al.*, (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins



the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ $\mu$ l) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-MeIm<sub>7</sub>), is then added to a final concentration of 10 mM 1-MeIm<sub>7</sub>. A ss DNA solution is then dispensed into CovaLink NH strips (75  $\mu$ l/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-MeIm<sub>7</sub>, is made fresh and 25  $\mu$ l added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995), 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res., 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) Anal. Biochem. 169(1), 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) Proc. Nat'l. Acad. Sci., USA 91(11), 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

#### 4.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24), 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*II, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *Cvi*II normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*Cvi*II\*\*), yield a quasi-random distribution of DNA fragments from the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *Cvi*II\*\* digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *Cvi*II\*\* restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 µg instead of 2-5 µg); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed).

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

#### 4.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm<sup>2</sup>, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient.

Where the 96 subarrays are identical, the dot span may be 1 mm<sup>2</sup> and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

## 5.0 EXAMPLES

### 5.1 EXAMPLE 1

#### Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences.

## 5.2 EXAMPLE 2

### Assemblage of Novel Contigs

The contigs of the present invention, designated as SEQ ID NO: 2083-2534 were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST, gb pri, and UniGene, and exons from public domain genomic sequences predicated by GenScan) that belong to this assemblage. The algorithm terminated when there were no additional sequences from the above databases that would extend the assemblage. Further, inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Table 8 sets forth the novel predicted polypeptides (including proteins) encoded by the novel polynucleotides (SEQ ID NO: 2083-2534) of the present invention, and their corresponding translation start and stop nucleotide locations to each of SEQ ID NO: 2083-2534. Table 8 also indicates the method by which the polypeptide was predicted. Method A refers to a polypeptide obtained by using a software program called FASTY (available from <http://fasta.bioch.virginia.edu>) which selects a polypeptide based on a comparison of the translated novel polynucleotide to known polynucleotides (W.R. Pearson, Methods in Enzymology, 183:63-98 (1990), herein incorporated by reference). Method B refers to a polypeptide obtained by using a software program called GenScan for human/vertebrate sequences (available from Stanford University, Office of Technology Licensing) that predicts the polypeptide based on a probabilistic model of gene structure/compositional properties (C. Burge and S. Karlin, J. Mol. Biol., 268:78-94 (1997), incorporated herein by reference). Method C refers to a polypeptide obtained by using a Hyseq proprietary software program that translates the novel polynucleotide and its complementary strand into six possible amino acid sequences (forward and reverse frames) and chooses the polypeptide with the longest open reading frame.

### 5.3 EXAMPLE 3

#### Novel Nucleic Acids

The novel nucleic acids of the present invention SEQ ID NO: 1-1041 were assembled from Hyseq 's proprietary EST sequences as described in Example 1 and human genome sequences that are available from the public databases (<http://www.ncbi.nlm.nih.gov/>).

Exons were predicted from human genome sequences using GenScan

(<http://genes.mit.edu/GENSCANinfo.html>); HMMgene

([http://www.cbs.dtu.dk/services/HMMgene/hmmgene1\\_1.html](http://www.cbs.dtu.dk/services/HMMgene/hmmgene1_1.html)); and GenMark.hmm

([http://genemark.biology.gatech.edu/GeneMark/whmm\\_info.html](http://genemark.biology.gatech.edu/GeneMark/whmm_info.html)). The Hyseq proprietary

EST sequences and the predicted exons were assembled based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%. Then, the predicted genes were analyzed using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark) for presence of a signal peptide. These sequences were further analyzed for absence of a

transmembrane region using the TMpred program

([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)).

Table 1 shows the various tissue sources of SEQ ID NO: 1-1041.

The homologs for polypeptides SEQ ID NO: 1042-2082, that correspond to nucleotide sequences SEQ ID NO: 1-1041 were obtained by a BLASTP version 2.0a1 19MP-WashU searches against Genpept release 124 using BLAST algorithm. The results showing homologues for SEQ ID NO: 1042-2082 from Genpept 124 are shown in Table 2.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6, 219-235 (1999), <http://motif.stanford.edu/ematrix-search/> herein incorporated by reference), all the polypeptide sequences were examined to determine whether they had identifiable signature regions. Scoring matrices of the eMatrix software package are derived from the BLOCKS, PRINTS, PFAM, PRODOM, and DOMO databases. Table 3 shows the accession number of the homologous eMatrix signature found in the indicated polypeptide sequence, its description, and the results obtained which include accession number subtype; raw score; p-value; and the position of signature in amino acid sequence.

Using the Pfam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4 shows the

name of the Pfam model found, the description, the e-value and the Pfam score for the identified model within the sequence. Further description of the Pfam models can be found at <http://pfam.wustl.edu/>.

The GeneAtlas™ software package (Molecular Simulations Inc. (MSI), San Diego, CA) was used to predict the three-dimensional structure models for the polypeptides encoded by SEQ ID NO 1-1041 (i.e. SEQ ID NO: 1042-2082). Models were generated by (1) PSI-BLAST which is a multiple alignment sequence profile-based searching developed by Altschul et al, (Nucl. Acids. Res. 25, 3389-3408 (1997)), (2) High Throughput Modeling (HTM) (Molecular Simulations Inc. (MSI) San Diego, CA,) which is an automated sequence and structure searching procedure (<http://www.msi.com/>), and (3) SeqFold™ which is a fold recognition method described by Fischer and Eisenberg (J. Mol. Biol. 209, 779-791 (1998)). This analysis was carried out, in part, by comparing the polypeptides of the invention with the known NMR (nuclear magnetic resonance) and x-ray crystal three-dimensional structures as templates. Table 5 shows: "PDB ID", the Protein DataBase (PDB) identifier given to template structure; "Chain ID", identifier of the subcomponent of the PDB template structure; "Compound Information", information of the PDB template structure and/or its subcomponents; "PDB Function Annotation" gives function of the PDB template as annotated by the PDB files (<http://www.rcsb.org/PDB/>); start and end amino acid position of the protein sequence aligned; PSI-BLAST score, the verify score, the SeqFold score, and the Potential(s) of Mean Force (PMF). The verify score is produced by GeneAtlas™ software (MSI), is based on Dr. Eisenberg's Profile-3D threading program developed in Dr. David Eisenberg's laboratory (US patent no. 5,436,850 and Luthy, Bowie, and Eisenberg, Nature, 356:83-85 (1992)) and a publication by R. Sanchez and A. Sali, Proc. Natl. Acad. Sci. USA, 95:13597-12502. The verify score produced by GeneAtlas normalizes the verify score for proteins with different lengths so that a unified cutoff can be used to select good models as follows:

Verify score (normalized) = (raw score – 1/2 high score)/(1/2 high score)

The PFM score, produced by GeneAtlas™ software (MSI), is a composite scoring function that depends in part on the compactness of the model, sequence identity in the alignment used to build the model, pairwise and surface mean force potentials (MFP). As given in table 5, a verify score between 0 to 1.0, with 1 being the best, represents a good

model. Similarly, a PMF score between 0 to 1.0, with 1 being the best, represents a good model. A SeqFold™ score of more than 50 is considered significant. A good model may also be determined by one of skill in the art based all the information in Table 5 taken in totality.

5           Table 6 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, 10 Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et al reference, was obtained for the polypeptide sequences.

15           Table 7 correlates each of SEQ ID NO: 1-1041 to a specific chromosomal location.

          Table 9 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-1041, their corresponding polypeptide sequences SEQ ID NO: 1042-2082, their corresponding priority contig nucleotide sequences SEQ ID NO: 2083-2534, their corresponding priority contig polypeptide sequences SEQ ID NO: 2535-2986, and the US 20 serial number of the priority application in which the contig sequence was filed.

          Table 10 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-1041, the novel polypeptide sequences SEQ ID NO: 1042-2082, and the corresponding SEQ ID NO in which the sequence was filed in priority US application 60/311,261.



114

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
adrenal gland	Clontech	ADR002	13 23 34 45 77 111 115 122 187 194 210-211 249-250 255 290 320 357-358 362 420 443 451 492 499 551 577 630 698 702 713 718 805 808 819 841-843 845 861 896 899 909 924 937 949 985 1037
adult bladder	Invitrogen	BLD001	9 87 189 320-321 358 563 768 840 970
adult brain	Clontech	ABR001	184-186 277 282 352 558 849 871 898 958
adult brain	Clontech	ABR006	30 45 170 199 210 226 260 292- 294 340 357 413 443-444 478 499 551-552 579 582 584-588 632-637 646 654-655 676 683 731-732 755-756 777 813-827 861 872 874 880 883 1002 1012
adult brain	Clontech	ABR008	15 45 54 61 67 81 87 101 106 108 122-123 143-144 170 181- 183 195-209 215 222 245-248 261-270 283-289 292-293 296 306 308-310 327 340 358 370 394-407 409 421 428 440 442 459 477-478 496 531-547 551- 552 556 565-566 578-579 606 618 620-621 629-630 651 653- 655 664 667-668 707 713-714 729 745 750 753 756 772 779 788 790 793-794 799-800 802 808 812 823 826-827 849-850 859 862 872 883 885 898 917 919 921 930 935-936 947 974 985-986 992 1002 1006 1012 1028 1030 1036 1039
adult brain	Clontech	ABR011	1012
adult brain	GIBCO	AB3001	23 57-58 67 85 296 492 499 579 853 898-899 950 1012
adult brain	GIBCO	ABD003	45 59-62 67 72 82 85-88 156 179-180 182 296 299 355-356 440 458 474 483 499 563 823 840 852 860 885 898 992 999 1012
adult brain	Invitrogen	ABR014	45 115 238 470 599 653 974-976
adult brain	Invitrogen	ABR015	45 600 885 1012
adult brain	Invitrogen	ABR016	599 1012
adult brain	Invitrogen	ABT004	34 45 54 74 84 118 138-143 170- 171 180-181 208 255 277 359 379 428 438 499 501 536 715 731 783 793 799 805 809 824 862 898 912 977 998 1012
adult cervix	BioChain	CVX001	23 26 48 54 57 67 77 118 121 177 183 238 255 271-272 296 303 311-319 325 352 361-362 411-412 419-420 424 428 440 447 478 541 567 569 599-600 622 699 793 805 813 831 836- 837 839 844-845 848 863 872

115

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
			913 928-929 944 958 965 970 973 1001 1004
adult colon	Invitrogen	CLN001	250 322-325 429 630 788 970 985
adult heart	GIBCO	AHR001	28-30 45 61 67 90-94 118 122 150-151 183 193 250-251 279 349-351 369-370 410 419 474 483 485 490 493 552 563 719 773 835-836 853 861 961 976 1030
adult kidney	GIBCO	AKD001	24 31-34 44-46 48 55 62 67 81 121 144 151 162 176-178 183 251 255 258 277 352 358 369- 370 386 408 420 429 483 490 536 546 579 599-600 602 645 698 793 805 874 898 913
adult kidney	Invitrogen	AKT002	32 53-54 67 85 177 251 260 341 386 408 419-420 431-436 478 490 493 507 561 582 596-599 698 728 788 805 819 837 844- 848 885 898 969 989 1013
adult liver	Clontech	ALV003	101 121 193 579 638-639 729 890-893 919 1007 1017
adult liver	Invitrogen	ALV002	75 157 173 183 212-214 236 240 263 292 323 335 386 408 415 495-499 552 577 589 599 727 782 858 869 898-900 924 968
adult lung	GIBCO	ALG001	67 77 152 369 386 419 443 483 583 732 849 907
adult ovary	Invitrogen	AOV001	5 26 34 43 45 48 55 61-62 64-67 77 87 101-102 105 115 118 122- 129 143 151 155-163 170 174- 175 177 181-183 193 251-252 286 292 338 347 353-354 369 381 410 415 420 424 451 458 483 489 497 499 515 536 541 546 552 577 579 595 599-600 604 647 658 661 665 699 744 782-783 800 805-806 814 831 835 839-840 844 853 874 895 898-899 913 924 929 941-942 949 973 977 994 1004 1007 1012 1016 1031 1037
adult placenta	Clontech	APL001	67 419 688 728 848 930
adult spleen	Clontech	SPLc01	82 101 187 255 260 358 370 447 483 489 579 586 648 768 835 845 848 853-857 863 885 913 917 962 986
adult spleen	GIBCO	ASP001	87 105 108 122 158 172 215 299 380 492 499 552 599 622 785 830 840 850 889
adult testis	GIBCO	ATS001	68-69 106 183 251 301 360 386 520 541 570 753 788 832 840 890 916
bone marrow	Clontech	BMD001	10-12 16-19 24-26 35 46 48 58 77 85 95-96 98-99 122 156 164

116

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
			172 187 222 251 385 424 429 458 478 483 489 519 568-569 599 622-623 630-631 696 700 758 765 794 844 914 919 924 944 971 985 992 1001 1017
bone marrow	GF	BMD002	23 45 81-82 104-105 115 136 144 156 170 172-173 181 183 247 287 292 306 319-320 327 362 370 418 478-483 489 492 536 548-552 565 569-570 572 579 596 599 614-622 630 640- 641 643 653 668 691 699 708 715-718 726 743 756 758 772 789 841 889 917 920 947 958 994 1006 1010 1037 1039
cultured preadipocytes	Stratagene	ADP001	121 255 400 490-494 511 629 689 758 793 835 861 913 944 949 984
endothelial cells	Stratagene	EDT001	34 45 54 58 67 120-122 144 151- 154 183 193 299 385 440 451 458 483 490 499 515 552 563 569 577 579 599 622-623 752 793 800 844-845 898-899 942 944 949
fetal brain	Clontech	FBR001	139 168 356 599 702 712 831 845 850 872-873 898 921 1037
fetal brain	Clontech	FBR004	138 168 250 363 873-875 882
fetal brain	Clontech	FBR006	14 29 45 51 81 87 101 104 118 131 143-144 157 171 177 206 208-209 215 229 238 251 261 273 279 283 291-293 326-332 358 362 370-371 397 400 402 413 419 428 461 472 485 551- 560 568-569 579 618 620 629- 630 653-657 659-661 663-673 675 700 714 739-742 744-746 766 779 793 809 815 819 822 840 850 859 862 872 875-885 930 958 972 995 1002 1006 1028 1030-1031 1038
fetal brain	GIBCO	HFB001	13-15 54-57 62 67 70-72 84 121 174 177 180 183 410 417 424 485 518 520 542 552 578-579 599 785 793 805 831-832 840 858 871 883 898-899 977 1012
fetal brain	Invitrogen	FBT002	7 45 49 144-149 157 180 255 263 356 493 501 600 630 707 748 832 845 858 913 1012
fetal heart	Invitrogen	FHR001	24 45 81-82 104 114-115 118 121 144 152 181 239 247 288 292 327 362 370 381 419 428 444 453 458 478 486 493 503 569 571 576 582 596 618 640 668 674-688 719-722 731 744 753 762 772 784 794 819 823 836 850 885 914 944 949 957- 958 1017

117

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
fetal kidney	Clontech	FKD001	82 107 208 458 483 485 536 758 760 819 836 894 1017
fetal kidney	Clontech	FKD002	61 101 105 183 189 238 247 263 292 327 340 370 405 416 419 517 569 586 620 648 668 689- 691 731 746-752 763 771-772 787-788 819 840 842 854 861 872 944 958 961 969
fetal kidney	Invitrogen	FKD007	116
fetal liver	Clontech	FLV002	410 429 454 692-695 704 781 805 894-895 1017
fetal liver	Clontech	FLV004	67 107 115 118 151 187 241 255 287 370 466 478 492 518 548 552 569 582 589 630 653 668 696-699 752-757 784 789 805 885 908 985
fetal liver	Invitrogen	FLV001	45 101 130-137 157 222 240 337 386 428-429 492 552 589 693 727 840
fetal liver-spleen	Columbia University	FLS001	1-9 18 20-23 27 34 36-38 45 55 67 70 83 89 94 118 122 158 164 172-173 177 183 219 238 240 246 251 292 299 323 335 338 358 369 376 385-386 397 408 416 419 421-422 429 451 456- 460 466 472 478 483 489-490 493 516 536 543 546 551 569- 573 579 586 588-589 593-595 599-603 619 622 668 676 691 699 702 724 731 734 743 787 789 794 800 805 834-835 840 848 853 874 880 885 890-891 899 908 910 923 926-927 930 939-940 944 949 958 973 980 992 999 1004 1007 1009 1013
fetal liver-spleen	Columbia University	FLS002	3 8 17 22 36-37 46 55 61 63 70 72 85 89-90 94 106 122 148 156 158 165 172 177 181 194 213 215 219 246 251 292 299 304- 307 323-324 338 346 355 366 371 374 380-381 386 392 397 410 417 421 440 455 462-464 466-468 489-490 492-493 507- 521 536 552 565-566 569 571- 576 592 596 599 619 630 650 655 661 688 698-699 712 718 723-729 731 735-737 753 767 783 824 831 834 840 845 871 885 891 894 899 902 906-909 913 923-930 940 943 949 958 973 980 992 999 1003 1007 1017 1032 1040-1041
fetal liver-spleen	Columbia University	FLS003	23 67 106 150 158 193 338 374 376 411 443 478 493 546 565 569-570 582 589 609-613 630 661 699 724 727-734 767 809 812 834-835 845 880 890 910

118

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
			929-930 958 973 980 985 1013
fetal lung	Clontech	FLG001	728 824 1008
fetal lung	Clontech	FLG004	115 668
fetal lung	Invitrogen	FLG003	120 183 322 333-336 476 516 691 831 835 850 1012
fetal muscle	Invitrogen	FMS001	45 338-339 365 369 386 429 431 496-497 789 793 856 970 1008 1019 1033 1035
fetal muscle	Invitrogen	FMS002	45 115 171 247 327 365 370 405 536 642-652 668 710-711 719 726 758-761 765 836 899 901 907 913 948 965 1037
fetal skin	Invitrogen	FSK001	29 57 67 74 81 118 152 177 180 193 294 340-342 345 375 397 419 437-443 445-451 454 475 532 541 546 565 598 604 630 650 668 728 742 772 789 793 804-805 823 828-830 837 840 849 899 901 922 958 970 1007 1022 1033
fetal skin	Invitrogen	FSK002	34 45 77 81 85 115 173 200 279 292-293 360 370 381 419 428- 429 451 466 490 551 569-570 579 600 604 630 647 668 698 700-706 729 731 746 750 758 762-766 768-773 780 794 840 850 859 861 885 901 911 913 957 961 965 973 1038
fibroblast	Stratagene	LFB001	55 72 143 255 490 502-505 587 599 627 861 863 885 984 1037
induced neuron-cells	Stratagene	NTD001	30 82 111 124 181 206 356 392 410 417 484-488 578 831-834 898 977 1036 1039
infant brain	Columbia University	IB2002	18 21 45 66 73-75 100-103 118 152 168-171, 177 180 241-242 252 292-295 340 345 366-367 413 438 454 499 501 542 561- 562 578-580 599 668 702 728- 729 745 765 768 772 793 796- 799 823-824 863 874 887 899 948-949 967 975 977 981 983 992 995 1012
infant brain	Columbia University	IB2003	81 101 113 118 177 180 241 252 293 340 345 367 371 379 381 400 417 499-501 536 562 578 580-581 629-630 702 713 745 796-805 824 831 837 840 845 874 885 967 977 981 985 1012 1030
infant brain	Columbia University	IBM002	168 358 413-414 913
infant brain	Columbia University	IBS001	415 417 533 581 886-888 977
leukocyte	Clontech	LUC003	77 619-889 949
leukocyte	GIBCO	LUC001	34 36 38-42 50-52 55 67 77 81- 83 85 121 137 144 158 172 183

119

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
			223 226 251 254 258 291 324 368-374 378 424 429 443 483 492 536 552 564 600 602 732 760 768 782 785 805 838 844- 845 848 850 889 898 905 908 946 973 992
lung	55 72 143 255 490 502-505 587 599 627 861 863 885 984 1037		
lung tumor	Invitrogen	LGT002	55 61 65 77-79 82 102 105 115 156-157 165-167 170 182-183 197 243-244 251 253 296-297 325 370 386 418-419 421-425 478 483 492 499 520 531 533 541 569 577 582 600 788 844- 845 848 874 899 911 913 916- 918 939 944 949 956 970 976
lymph node	Clontech	ALN001	47 63 104-105 183 483 492 691 894 1017
lymphocytes	ATCC	LPC001	45 53 77 158 193 251 392 421 455 469-474 483 507 536 546 579 581 618 621 640 765 780- 787 793 838 845 875 924 968 978 999
macrophage	Invitrogen	HMP001	122 147 157 183 251 255 493 738 898-899 903-905
mammary gland	Invitrogen	MMG001	45 64 67 83-84 101 113 143 148 152 158 164 177 181-183 189 216-218 253 255 258 263 274 299 336 419 421 423 426-430 440 466 478 490 520 533 536 564 569 579 582 630 646 753 768 782 789 800 835 840 848 850 883 912-913 944 950 958
melanoma from-cell-line- ATCC-#CRL-1424	Clontech	MBL004	62 158 181 298 362 364 402 419 515 536 896-897 958 973 1004 1008
*Mixture of 16 tissues - mRNA	Various Vendors	CGd010	353 358 823 942 982 1020
*Mixture of 16 tissues - mRNA	Various Vendors	CGd011	569 630 944 955 999
*Mixture of 16 tissues - mRNA	Various Vendors	CGd012	9 38 59 63 80 85 122-123 152 154 177 195 217 232 246 250 296 300 306 323-324 381 427 434 438-439 478 489 499 507 517 538 558 565 571 575 630 657 681 701 736 762 792 800 802 823-824 861 871-872 899 929 941 955 968 974 985-1003 1006 1011-1012 1033
*Mixture of 16 tissues - mRNA	Various Vendors	CGd013	232 434 748 956-958 992
*Mixture of 16 tissues - mRNA	Various Vendors	CGd015	18 69 115 324 335 548 551 569 582 600 622 731 819 899 911 944 957-958 1012 1017-1018

120

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
*Mixture of 16 tissues - mRNA	Various Vendors	CGd016	46 172 183 323 371 481 493 565 569 571 596 599 630 654 698 745 762 786 849 907 944 1004- 1013 1037 1039
neuronal cells	Stratagene	NTU001	7 33 45 107 113 121 150 183 286 385 440 478 483 485 487 489 536 569 582 756 768 772 819 836 944 958 966 1001
pituitary gland	Clontech	PIT004	158 222 255 345 356 370 379 569 579 819 831 861-862 885 898 922 1017
placenta	Clontech	PLA003	7 36 61 279 419 478 489 582 586 599 641 647 668 681 707-711 774-779 1001
placenta	Invitrogen	APL002	57 173 536 728 793 800
prostate	Clontech	PRT001	26 219-222 229 412 599 665 762 835 837 860 878 951 1031
rectum	Invitrogen	REC001	9 292 343-346 431 546 714 800 863 918
retinoic acid-induced-neuronal-cells	Stratagene	NTR001	112 400 478 569 582 629 756 758 800 819 831 835-836 850 906 944 958
salivary gland	Clontech	SAL001	58 61 77 118 150 158 294 347- 348 483 492-493 546 752 830 915
skeletal muscle	Clontech	SKM001	80 118 247 365 483 719 805 812 823
small intestine	Clontech	SIN001	34 37 45 52 60 93 106 119 121 138 144 177 180 208 223-225 238 247 294 323 335-336 343 362 370 380 386 397 409-411 416 420 440 451 455 478 489 493 536 571 577 579 590 602 604-608 614 622 624-628 655 668 688 700 714 805-812 831 841 872 894 899 914 924 926 929 958 961 965 973 991 998 1017
spinal cord	Clontech	SPC001	51 164 182-183 190 226-228 255-257 275-277 286 296 299 451 454 542 552 579 591 728 753 770 786 790 831 835 849- 852 898 907 958 1000 1012
stomach	Clontech	STO001	72 222 232 247 258 366 645
thalamus	Clontech	THA002	45 49 113 155 164 180 183 191- 192 208 229-232 238 345 417 443 512 551 558 592 630 728 800 823 840 858-860 885 898 976 1012
thymus	Clontech	THM001	45 141 160 183 258 360 378-379 418 451 460 569 602 619 731 788-790 819 835 845 958 965 1004
thymus	Clontech	THMc02	47 108 115 121 144 157 173 247 259-260 300 327 340 358 362 375-393 409 453 455 461 478-

121

Table 1

Tissue Origin	RNA/Tissue Source	Library Name	SEQ ID NO:
			479 489 551 565 569-570 579 582 615 630 640 653 668 708 744 752 758 766 790-795 810 819 823 835-836 845 850 853 861 885 911 919 938 958 962 994 1001 1027
thyroid gland	Clontech	THR001	46 58 67 80 82 144 160 177 183 193-194 233-235 251 255 263 268 278-280 286 299 301-303 324 358 370 386 397 408 410 420 440 474 483 493 506 519- 520 533 594 599-600 602 658 661 719 758 772 785 788 793 830 851 853 864-867 898 904 909 924 929 961 973 991 998 1001 1009
trachea	Clontech	TRC001	45 154 236 238 281 323 416 571 602 868-869 913
umbilical cord	BioChain	FUC001	34 45 54 58 67 70 85 152 154 177 180 188 208 251 299 370 409 415 419 434 451-455 483 596 599 647 661 733 742 793 808 839-840 845 849-850 861 888 911 913 992
uterus	Clontech	UTR001	177 237-239 255 258 417 493 520 567 599 604 646 844 870 874 898 973
young liver	GIBCO	ALV001	45 419 440 443 490 653 732 753 805 845 898 904

\*The 16 tissue/mRNAs and their vendor sources are as follows: 1) Normal adult brain mRNA (Invitrogen), 2) Normal adult kidney mRNA (Invitrogen), 3) Normal fetal brain mRNA (Invitrogen), 4) Normal adult liver mRNA (Invitrogen), 5) Normal fetal kidney mRNA (Invitrogen), 6) Normal fetal liver mRNA (Invitrogen), 7) normal fetal skin mRNA (Invitrogen), 8) human adrenal gland mRNA (Clontech), 9) Human bone marrow mRNA (Clontech), 10) Human leukemia lymphoblastic mRNA (Clontech), 11) Human thymus mRNA (Clontech), 12) human lymph node mRNA (Clontech), 13) human so\spinal cord mRNA (Clontech), 14) human thyroid mRNA (Clontech), 15) human esophagus mRNA (BioChain), 16) human conceptional umbilical cord mRNA (BioChain).



122

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1044	AAB32400	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 30 SEQ ID NO:86.	339	100
1044	AAM74711	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 35017.	335	100
1044	AAM61909	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 34014.	335	100
1045	gi3859599	Arabidopsis thaliana	similar to class I chitinases (Pfam: PF00182, E=1.2e-142, N=1)	74	27
1045	gi15292107	Drosophila melanogaster	LD38671p	74	33
1045	gi2258324	Fusarium oxysporum f. sp. ciceris	yellowing-associated protein	73	32
1046	gi17428204	Ralstonia solanacearum	CONSERVED HYPOTHETICAL PROTEIN	74	32
1046	gi4314432	Homo sapiens	similar to phosphatidylinositol (4,5)bisphosphate 5-phosphatase; match to PID:gi1399105	71	30
1046	gi 17545909 ref NP_519311.1	Ralstonia solanacearum	CONSERVED HYPOTHETICAL PROTEIN	74	32
1047	gi9756017	Actinoplanes sp. 50/110	alpha-amylase	69	38
1047	gi 6572499 gb AAF17291.1	Homo sapiens	LHX3 protein	67	26
1047	gi 18572988 ref XP_029170.2	Homo sapiens	LIM homeobox protein 3	67	26
1048	AAAY28474	Homo sapiens	UYJO Human Capon protein.	721	99
1048	gi2895555	Homo sapiens	carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase	721	99
1048	gi2895557	Rattus norvegicus	carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase	654	92
1049	gi19713721	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	GTP-binding protein era	66	28
1050	gi31291	Homo sapiens	fumarylacetoacetase (AA 1-349)	175	70
1050	gi182393	Homo sapiens	fumarylacetoacetate hydrolase	175	70
1050	gi12803409	Homo sapiens	fumarylacetoacetate	175	70
1052	gi4680089	Human immunodeficiency virus type 1	envelope glycoprotein	79	26
1052	gi3868997	Ephydatia fluviatilis	BFPDE2	74	20
1052	gi4679590	Human immunodeficiency virus type 1	envelope glycoprotein	74	25
1054	gi3844648	Mycoplasma genitalium	glycerol kinase (glpK)	71	28

123

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1054	gi18448155	Ipomoea leaf curl virus	AC3	70	27
1054	gi 12044888 ref NP_072698.1	Mycoplasma genitalium	glycerol kinase (glpK)	71	28
1056	AAM56747	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 28852.	229	72
1056	AAM67067	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27373.	224	69
1056	AAM54664	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26769.	224	69
1058	gi 13310191 gb AAK18189.1 AF331500_1	multiple sclerosis associated retrovirus element	recombinant envelope protein	228	79
1058	gi 21103962 gb AAM33141.1	Homo sapiens	enverin-2	209	77
1058	gi 8272468 gb AAF74215.1 AF156963_1	Homo sapiens	envelope protein	198	75
1059	gi20380199	Homo sapiens	Similar to LOC168246	251	100
1059	gi 8388692 emb CAB94042.1	Leishmania major	probable DNA-binding protein	67	46
1060	gi 21292780 gb EAA04925.1	Anopheles gambiae str. PEST	agCP4203	70	39
1061	gi330862	Equine herpesvirus 1	membrane glycoprotein	179	30
1061	gi17221106	Equine herpesvirus 1	glycoprotein gp2	178	34
1061	AAE03643	Homo sapiens	INCY- Human extracellular matrix and cell adhesion molecule-7 (XMAD-7).	175	29
1062	gi 11037117 gb AAG27485.1 AF194537_1	Homo sapiens	NAG13	334	66
1062	gi 1335205 emb CAA36480.1	Homo sapiens	ORFII	332	66
1063	gi21323402	Corynebacterium glutamicum ATCC 13032	ABC-type transporter, periplasmic component	70	36
1063	gi 19551869 ref NP_599871.1	Corynebacterium glutamicum	COG1464:ABC-type uncharacterized transport systems, periplasmic component	70	36
1063	gi 17551878 ref NP_4990	Caenorhabditis elegans	TPR Domain	67	37

124

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	90.1]				
1064	gi2308977	Aspergillus nidulans	chitin synthase	66	29
1065	gi18076958	Yarrowia lipolytica	Opt1 protein	74	30
1065	gi786145	Walleye dermal sarcoma virus	envelope polypeptide	73	28
1065	gi2801522	Walleye dermal sarcoma virus	gPr env	73	28
1066	gi9294279	Arabidopsis thaliana	Ta11-like non-LTR retroelement protein-like; CHP-rich zinc finger protein-like	67	32
1066	gi20848817 refXP_138010.1]	Mus musculus	similar to HEAT SHOCK COGNATE PROTEIN 80	83	69
1069	AAM77637	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 37943.	96	65
1069	AAM64901	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 37006.	96	65
1069	gi17473741 refXP_062380.1]	Homo sapiens	similar to Meningioma-expressed antigen 6/11 (MEA6) (MEA11)	112	56
1070	gi296288	Homo sapiens	histone H1	77	44
1070	gi5923857	Artemisia annua	squalene synthase	75	35
1070	AAO08837	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 22729.	73	39
1071	gi21483554	Drosophila melanogaster	SD02058p	72	29
1071	gi8515845	Homo sapiens	hepatocellular carcinoma associated protein TD26	71	38
1071	gi21483554 gb AAM52752.1]	Drosophila melanogaster	SD02058p	72	29
1072	gi5902896	Streptomyces avermitilis	type I polyketide synthase AVES 4	74	50
1072	gi21301752 gb EAA13897.1]	Anopheles gambiae str. PEST	agCP8235	70	34
1073	AAV30916_aa1	Homo sapiens	GEMY Human secreted protein AR415_4 cDNA.	99	66
1073	ABB89113	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1489.	99	66
1073	AAB90679	Homo sapiens	GEMY Human AR415_4 protein sequence SEQ ID 35.	99	66
1074	AAG99338	Homo sapiens	TAKE Human atypical tachykinin protein fragment SEQ ID NO: 20.	380	92
1074	AAG99336	Homo sapiens	TAKE Human atypical tachykinin protein fragment SEQ ID NO: 13.	329	91
1074	AAG99333	Homo sapiens	TAKE Human atypical tachykinin protein fragment SEQ ID NO: 3.	324	91
1075	gi17945760	Drosophila melanogaster	RE33302p	305	29

125

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1075	gi1039447	Saccharomyces cerevisiae	Lpb1p	91	25
1075	AAB64777	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 5 SEQ ID NO:63.	78	77
1076	AAB50261	Homo sapiens	CORI- Human breast cancer associated B726P-20 protein.	308	39
1076	AAB50244	Homo sapiens	CORI- Human breast cancer associated B726P-79 protein.	308	39
1076	AAB84702	Homo sapiens	CORR Amino acid sequence of a human cancer associated antigen.	308	39
1077	gi2529735	Gorilla gorilla	glycophorin B/E precursor	71	31
1077	AAB74724	Homo sapiens	INCY- Human membrane associated protein MEMAP-30.	70	31
1077	gi4164424	Schizosaccharomyces pombe	similar to yeast cytoskeleton control protein Bni1p	70	24
1078	gi18145107	Clostridium perfringens	probable transcriptional regulator	71	28
1078	gi 9581801 emb CAC00546.1	Plasmodium falciparum	guanylyl cyclase	69	24
1078	gi 16805032 ref NP_473061.1	Plasmodium falciparum	Ser/Thr protein kinase	69	26
1079	gi 20886321 ref XP_140614.1	Mus musculus	similar to olfactory receptor, family 5, subfamily V, member 1; olfactory receptor, family 5, subfamily V member 1	72	34
1081	gi9650824	Petroselinum crispum	common plant regulatory factor 5	76	28
1081	gi559695	Hydrolyagus collicii	This CDS feature is included to show the translation of the corresponding C_region. Presently translation qualifiers on C_region features are illegal	74	31
1081	gi476622	Hydrolyagus collicii	immunoglobulin light chain	74	31
1082	AAM39205	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2350.	363	71
1082	AAO07159	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 21051.	357	76
1082	AAM40991	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5922.	343	79
1083	gi 17229222 ref NP_485770.1	Nostoc sp. PCC 7120	similar to HetF protein	72	30
1084	gi17221628	Felis catus	T-lymphocyte surface CD2 antigen	76	38
1084	gi18565073	Crimcan-Congo hemorrhagic fever virus	envelope glycoprotein precursor	74	29
1084	gi 17221628 dbj BAB78475.1	Felis catus	T-lymphocyte surface CD2 antigen	76	38
1085	gi17430213	Ralstonia	PUTATIVE HEMAGGLUTININ-	74	26

126

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		solanacearum	RELATED PROTEIN		
1087	gi2323287	multiple sclerosis associated retrovirus	polyprotein	618	79
1087	gi4996596 dbj BAA7854.9.1	Human endogenous retrovirus W	polyprotein	317	74
1087	gi9630708 ref NP_047255.1	Feline leukemia virus	gag-pol precursor polyprotein gPr80	293	38
1088	gi15075953	Sinorhizobium meliloti	PUTATIVE MOLYBDENUM TRANSPORT SYSTEM PERMEASE ABC TRANSPORTER PROTEIN	70	56
1088	gi2288880	Arthrobacter nicotinovorans	transmembrane protein	67	56
1088	gi17298547	Bradyrhizobium japonicum	ModB	67	56
1089	AAU95660	Homo sapiens	ZYMO Human Znt2 protein.	231	61
1089	AAU83682	Homo sapiens	GETH Human PRO protein, Seq ID No 182.	210	59
1089	AAU99386	Homo sapiens	GETH Human PRO1305 (UNQ671) amino acid sequence SEQ ID NO:153.	210	59
1090	gi7688355	Solanum tuberosum	Dof zinc finger protein	70	31
1090	gi4389445	Drosophila melanogaster	transcription factor	67	32
1090	gi7688355 emb CAB89831.1	Solanum tuberosum	Dof zinc finger protein	70	31
1092	AAG78884	Homo sapiens	BIOW- Human ribosomal protein s5-17.	90	44
1092	AAM91239	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:18832.	72	53
1092	AAM95026	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 3684.	72	48
1094	gi18676450	Homo sapiens	FLJ00122 protein	69	38
1094	gi18073428	Homo sapiens	stabilin-2	69	38
1094	gi20806091 ref NP_060034.8	Homo sapiens	stabilin-2; CD44-like precursor FELL	69	38
1095	gi20906397	Methanosarcina mazei Goel	conserved protein	76	44
1095	gi21299784 gb EAA11929.1	Anopheles gambiae str. PEST	agCP6531	75	30
1095	gi17549046 ref NP_522386.1	Ralstonia solanacearum	CONSERVED HYPOTHETICAL PROTEIN	73	32
1096	AAB58317	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 655.	678	100
1096	gi862600	Drosophila melanogaster	male-specific lethal-1 protein	176	25

127

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1096	gi601930	Oryctolagus cuniculus	neurofilament-H	115	24
1097	AAU83109	Homo sapiens	ZYMO Novel secreted protein Z701935G4P.	76	85
1097	gi 20348496 ref XP_111712.1	Mus musculus	similar to RIKEN cDNA 9030605E16	72	57
1098	gi18031887	Mus musculus	Fanconi anemia complementation group G	77	29
1098	gi12002137	Mus musculus	Fanconi anemia group G protein	77	29
1098	AAB72381	Homo sapiens	LEEM/ Human hairy and enhancer of Split homologue amino acid sequence.	75	28
1099	gi8217648	Homo sapiens	dJ579F20.1 (high-mobility group (nonhistone chromosomal) protein 1-like 1)	159	70
1099	gi5815432	Gallus gallus	high mobility group protein HMG1	154	70
1099	gi4140289	Gallus gallus	high mobility group 1 protein	154	70
1100	ABB11527	Homo sapiens	HYSE- Human apolipoprotein B receptor homologue, SEQ ID NO:1897.	84	26
1100	gi487347	Homo sapiens	breakpoint cluster region protein	81	32
1100	gi144050	Bordetella pertussis	filamentous hemagglutinin	78	30
1102	AAM68946	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29252.	327	81
1102	AAM79768	Homo sapiens	HYSE- Human protein SEQ ID NO 3414.	324	80
1102	AAM78784	Homo sapiens	HYSE- Human protein SEQ ID NO 1446.	324	80
1103	AAZ11186_aal	Homo sapiens	SAGA Gene encoding transmembrane domain containing protein clone HP02239.	143	68
1103	AAD31079_aal	Homo sapiens	INCY- Human cornichon protein (CORN) cDNA.	143	68
1103	AAA88439_aal	Homo sapiens	GETH Antitumour PRO181 cDNA clone DNA23330-1390.	143	68
1104	ABB07527	Homo sapiens	INCY- Human drug metabolizing enzyme (DME) (ID: 5643401CD1).	562	100
1104	ABB07515	Homo sapiens	INCY- Human drug metabolizing enzyme (DME) (ID: 8097779CD1).	562	100
1104	gi13161409	Mus musculus	family 4 cytochrome P450	431	76
1107	gi13542874	Mus musculus	Similar to CGI-67 protein	677	64
1107	AAU81978	Homo sapiens	INCY- Human secreted protein SECP4.	665	65
1107	AAU77137	Homo sapiens	MILL- Human alpha/beta hydrolase 38618 polypeptide.	665	65
1108	gi13620885	Homo sapiens	mitochondrial ribosomal protein S6	323	100
1108	gi13620887	Mus musculus	mitochondrial ribosomal protein S6	284	82
1108	gi19713140	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	Fusobacterium outer membrane protein family	79	28
1109	gi18378673	Homo sapiens	PATE	607	89
1109	gi5305193	Rattus norvegicus	sperm protein 10	108	30

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1109	gi969103	Mus musculus	mSP-10	107	27
1110	gi2462979	Bos taurus	Tenascin-X	119	34
1110	gi3413958	Homo sapiens	LDL receptor related protein 105	110	27
1110	gi13938519	Homo sapiens	low density lipoprotein receptor-related protein 3	110	27
1111	gi17981053	Mus musculus	transcription factor NFAT5	82	32
1111	gi15425825	Mus musculus	tonicity-responsive enhancer binding protein	82	32
1111	gi6911148	Mus musculus	transcription factor NFAT5 isoform b	82	32
1112	gi6634473	Metarhizium anisopliae var. anisopliae	adenylate cyclase, ACY	73	30
1113	AAU19759	Homo sapiens	HUMA- Human novel extracellular matrix protein, Seq ID No 409.	900	70
1113	gi3171934	Mus musculus	neuronal-STOP protein	886	52
1113	gi2769587	Mus musculus	STOP protein	885	52
1114	gi18652188	Oenococcus oeni	OppF	72	41
1115	gi9119	Drosophila sp.	fos-related antigen	69	37
1115	gi7769652	Drosophila melanogaster	Fos-related antigen	69	37
1115	gi17862946	Drosophila melanogaster	SD04477p	69	37
1116	gi21212948	Mus musculus	peroxisomal protein (PeP)	243	83
1116	gi2347114	Mus musculus	CC chemokine receptor-5	72	28
1116	gi2431976	Mus musculus	CCR5	72	28
1117	gi20825251 ref XP_131998.1	Mus musculus	similar to RE1-silencing transcription factor; neuron restrictive silencer factor; repressor binding to the X2 box	77	40
1117	gi15597871 ref NP_251365.1	Pseudomonas aeruginosa	probable type II secretion system protein	69	41
1118	gi3860513 emb CAA13574.1	Mus famulus	reverse transcriptase	303	82
1118	gi3860536 emb CAA13577.1	Mus saxicola	reverse transcriptase	303	81
1118	gi3860510 emb CAA13573.1	Mus dunni	reverse transcriptase	298	63
1119	AAO04758	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 18650.	234	59
1119	AAM69569	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29875.	220	63
1119	AAM67717	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 28023.	219	49
1120	gi21107877	Xanthomonas axonopodis pv. citri str. 306	cytochrome C	78	27
1120	gi15292331	Drosophila melanogaster	LD47230p	77	42
1120	gi15072444	Avian	phosphoprotein	72	38

129

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		paramyxovirus 6			
1121	AAB44126	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1571.	150	83
1121	gi550015	Homo sapiens	ribosomal protein L21	150	83
1121	gi619788	Homo sapiens	L21 ribosomal protein	150	83
1122	AAU74448	Homo sapiens	OULU- Human protein sequence of lysyl hydroxylase 1 (LH1).	125	100
1122	gi190074	Homo sapiens	lysyl hydroxylase	125	100
1122	gi5817297	Homo sapiens	lysyl hydroxylase 1	125	100
1123	gi21281601	Caenorhabditis elegans	C. elegans PQN-44 protein (corresponding sequence F55A12.9c)	78	34
1123	gi14578225	Caenorhabditis elegans	C. elegans PQN-44 protein (corresponding sequence F55A12.9b)	76	38
1123	gi2088669	Caenorhabditis elegans	C. elegans PQN-44 protein (corresponding sequence F55A12.9a)	76	38
1125	AAU17301	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 866.	344	88
1125	AAE11776	Homo sapiens	INCY- Human kinase (PKIN)-10 protein.	344	88
1125	AAU17304	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 869.	340	86
1126	AAM41712	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6643.	152	96
1126	AAM39926	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3071.	152	96
1126	AAM79067	Homo sapiens	HYSE- Human protein SEQ ID NO 1729.	152	96
1127	AAE02938	Homo sapiens	MILL- Human adenylate cyclase 25678.	252	98
1127	AAB02006	Homo sapiens	TEXA Adenylyl cyclase type II-C2 C2 alpha domain.	252	98
1127	gi202752	Rattus norvegicus	adenylyl cyclase type II	252	98
1128	AAA94860_aal	Homo sapiens	TEXA Human caspase activator Smac coding sequence.	96	100
1128	AAU78447	Homo sapiens	UYJE- Inhibitor of apoptosis (IAP) protein Smac.	96	100
1128	AAB26210	Homo sapiens	TEXA Human caspase activator Smac.	96	100
1129	gi3874765	Caenorhabditis elegans	Similarity to Drosophila acetylcholine receptor protein (SW:ACH1_DROME), contains similarity to Pfam domain: PF00065 (Neurotransmitter-gated ion-channel), Score=296.9, E-value=5e-86, N=3	97	30
1129	gi6681597	Yaba monkey tumor virus	similar to vaccinia G8R	72	28
1129	gi 17548199 ref NP_509932.1	Caenorhabditis elegans	acetylcholine receptor	97	30
1130	gi 17564116 ref NP_506484.1	Caenorhabditis elegans	tyrosine-protein kinase	73	29
1131	gi13925613	Homo sapiens	insulinoma-associated protein IA-6	88	27
1131	gi158485	Drosophila	son of sevenless protein	85	24



130  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		melanogaster			
1131	gi7287782	05-Feb-1998	symbol=Sos; synonym=BG:DS00941.4; match=method:"sim4", score:"1000.0", desc:"GenBank::M83931:Drosophila melanogaster son of sevenless (Sos) mRNA, complete cds. CDS:346..5133; PID:g158485.", species:"Drosophila melanogaster"; match=method:"BLASTX", version:"2.0a19MP-WashU [Build sol2.5-ultra 01:47:30	85	24
1132	gi9696	Mytilus edulis	polyphenolic adhesive protein	75	25
1134	gi13562016	Plectreurys tristis	fibroin 2	72	29
1134	gi1129074	Bacillus subtilis	beta-N-acetylglucosaminidase	69	28
1134	gi2636104	Bacillus subtilis	N-acetylglucosaminidase (major autolysin) (CWBP90)	69	28
1135	AAB58870	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 578.	72	80
1135	gi11595476	Homo sapiens	RPB11b1beta protein	72	80
1135	AAB44840	Homo sapiens	HUMA- Human secreted protein encoded by gene 11.	69	45
1137	gi206985	Rattus norvegicus	troponin I	70	46
1137	gi16945895	Takifugu rubripes	SUN-like 1	70	31
1137	gi18394466 ref NP_058881.1	Rattus norvegicus	troponin I, skeletal, fast 2	70	46
1140	AAO04998	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 18890.	277	96
1140	gi19917538	Methanosarcina acetivorans str. C2A] [Methanosarcina acetivorans C2A	mttA/Hcf106 protein	80	28
1140	gi4959705	Mus musculus	fibulin-2	76	28
1141	gi10141010	Vesicular exanthema of swine virus	non-structural polyprotein	91	31
1141	gi6566147	Drosophila melanogaster	large Forked protein	85	30
1141	gi2317953	murid herpesvirus 4	glycoprotein 150	79	28
1142	AAB54067	Homo sapiens	HUMA- Human pancreatic cancer antigen protein sequence SEQ ID NO:519.	218	56
1142	gi1710365	Mus musculus	noggin	89	29
1142	gi21105761	Equus caballus	noggin	89	29
1143	gi21295753 gb EAA07898.1	Anopheles gambiae str. PEST	agCP1560	69	26
1144	gi505094	Homo sapiens	similar to an actin bundling protein,	127	35

131

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			dematn.		
1144	gi2337952	Homo sapiens	actin-binding double-zinc-finger protein	122	36
1144	gi21304227	Oryza sativa	ovule development aintegumenta-like protein BNM3	76	29
1145	gi 21298336 gb EAA10481.1	Anopheles gambiae str. PEST	agCP2121	68	37
1146	AAW22049	Homo sapiens	INCY- Interferon gamma inducing factor-2 (IGIF-2) alternate transcript variant.	221	100
1146	AAV05368_aa1	Homo sapiens	SCHE cDNA encoding human interleukin-1-gamma.	167	84
1146	AAH78060_aa1	Homo sapiens	STRD Nucleotide sequence of human interleukin 18 (IL-18).	167	84
1147	AAV57937	Homo sapiens	INCY- Human transmembrane protein HTPMPN-61.	123	100
1147	gi 20345904 ref XP_109823.1	Mus musculus	similar to delta-like homolog (Drosophila)	105	86
1148	gi19069293	Encephalitozoon cuniculi	similarity to ADP/ATP CARRIER PROTEIN	75	32
1148	gi8978336	Arabidopsis thaliana	contains similarity to CHP-rich zinc finger protein-gene_id:K23F3.4	74	26
1148	gi19716318	Aspergillus flavus	antigenic cell wall protein MP1	74	32
1149	gi5456699	Emericella nidulans	ATP-binding cassette multidrug transport protein ATRC	70	35
1149	gi 20898840 ref XP_139387.1	Mus musculus	similar to HSPC038 protein	69	31
1150	gi3883128	Arabidopsis thaliana	arabinogalactan-protein	96	32
1150	gi17429208	Ralstonia solanacearum	CONSERVED HYPOTHETICAL PROTEIN	92	26
1150	gi4063766	Emericella nidulans	chitinase	91	27
1151	gi13561058	Homo sapiens	dJ1108D11.1 (novel protein similar to C. elegans T22C1.7)	107	31
1151	gi21105299	Mytilus galloprovincialis	precollagen-NG	105	26
1151	gi14164347	Oncorhynchus mykiss	collagen a1(I)	96	28
1152	gi18479434	Mus musculus	olfactory receptor MOR188-1	76	33
1152	gi2653915	Oran virus	glycoprotein G1 and G2 precursor; envelope glycoprotein precursor	72	46
1152	gi18479436	Mus musculus	olfactory receptor MOR188-2	72	33
1153	gi3403167	Homo sapiens	GBAS	161	86
1153	gi12804791	Homo sapiens	glioblastoma amplified sequence	161	86
1153	AAB57149	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1727.	134	81
1154	gi17742234	Agrobacterium tumefaciens str. C58 (U.	histidase	87	35

132  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		Washington)			
1154	gi15159496	Agrobacterium tumefaciens str. C58 (Cereon)	AGR_L_1400Gmp	87	35
1154	gi158521	Drosophila melanogaster	seven-up protein type 2	80	32
1155	gi 10441551 gb AAG17099.1 AF189115.1	Cryptotermes domesticus	cytochrome b	65	28
1156	AAO12089	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 25981.	475	98
1156	gi20147787	Xenopus laevis	nuclear receptor corepressor	74	25
1156	gi19881705	Oryza sativa	Putative transposable element	72	32
1157	gi9963851	Homo sapiens	HT019	80	34
1157	AAB93530	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12884.	77	34
1157	gi1040970	Homo sapiens	fus-like protein	77	42
1158	gi9795254	Sepia officinalis	GABA-A receptor beta subunit	71	27
1158	gi15026157	Clostridium acetobutylicum	amidase, germination specific (cwIC/cwID B.subtilis ortholog)	68	34
1158	gi 9795254 gb AAF97816.1	Sepia officinalis	GABA-A receptor beta subunit	71	27
1159	AAB93423	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12641.	336	100
1159	gi13097768	Homo sapiens	Similar to RIKEN cDNA 2900073H19 gene	336	100
1159	gi20071708	Mus musculus	RIKEN cDNA 2900073H19 gene	334	96
1160	AAM72558	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 32864.	274	100
1160	AAM59959	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 32064.	274	100
1161	AAB07704	Homo sapiens	INMR Protein encoded by the endogenetic fragment of HERV-W.	139	36
1161	gi8272464	Homo sapiens	gag	139	36
1161	gi 5726238 gb AAD48375.1 AF123881.1	multiple sclerosis associated retrovirus element	gag polyprotein	131	35
1162	AAU25448	Homo sapiens	INCY- Human mddt protein from clone LG:1083264.1:2000MAY19.	346	79
1162	AAU11265	Homo sapiens	BODE- Human zinc finger protein 51.	319	65
1162	AAB95637	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18371.	314	67
1163	gi14189950	Homo sapiens	connexin 58	536	84
1163	gi9957542	Homo sapiens	connexin 59	536	84
1163	gi10946367	Danio rerio	connexin 55.5	485	81
1164	gi755700	Bombyx mori	sericin1B	76	27
1164	gi19569861	Dictyostelium discoideum	RTOA protein (Ratio-A).	76	28

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1164	gi10580635	Halobacterium sp. NRC-1	Vng1087c	76	25
1165	gi19915386	Methanosarcina acetivorans str. C2A] [Methanosarcina acetivorans C2A	WD-domain containing protein	89	28
1165	gi5639663	Homo sapiens	WD repeat protein WDR3	83	28
1165	gi11544739	Homo sapiens	dJ776P7.2 (WD repeat domain 3)	83	28
1166	AAM69338	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29644.	72	31
1166	AAM56953	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 29058.	72	31
1166	gi20197507	Arabidopsis thaliana	expressed protein	67	39
1167	gi5802812	Homo sapiens	Gag protein	83	30
1167	gi7160650	Bordetella bronchiseptica	pertactin (P.68)	79	31
1167	gi13173444	Bordetella bronchiseptica	pertactin	79	31
1168	gi1495029	Danio rerio	protein kinase CK2 alpha'	84	24
1168	gi643443	Penicillium chrysogenum	PHOG	82	32
1168	gi18858419 ref NP_571315.1	Danio rerio	casein kinase 2 alpha 2	84	24
1169	gi206716	Rattus norvegicus	salivary proline-rich protein	90	31
1169	gi15029903	Mus musculus	Similar to proline-rich protein BstNI subfamily 2	89	36
1169	gi53182	Mus musculus	proline rich protein	81	34
1170	gi17553370 ref NP_498318.1	Caenorhabditis elegans	F40H6.5.p	78	33
1170	gi15215731 gb AAK91411.1	Arabidopsis thaliana	AT4g36780/C7A10_580	73	30
1171	gi340446	Homo sapiens	zinc finger protein 7 (ZFP7)	218	61
1171	AAB43928	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1373.	216	58
1171	AAB21040	Homo sapiens	INCY- Human nucleic acid-binding protein, NuABP-44.	213	48
1172	AAE04368	Homo sapiens	INCY- Human kinase (PKIN)-9.	120	85
1172	AAM79153	Homo sapiens	HYSE- Human protein SEQ ID NO 1815.	120	85
1172	AAE10614	Homo sapiens	CURA- Human novel STE20-like protein, NOV-3d.	120	85
1173	gi218572	Pan troglodytes	prot GOR	74	29
1173	gi243898	Pan	GOR	74	29
1173	gi1666473	Mus musculus	NOV protein	71	50
1174	gi5901830	Drosophila melanogaster	BcDNA.GH07910	74	31

134

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1174	AAM80237	Homo sapiens	HYSE- Human protein SEQ ID NO 3883.	71	38
1174	ABB11528	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:1898.	71	38
1175	gi 12054759 emb CAC20748.1	Podospora anserina	catalase A	65	33
1176	AAM93289	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2777.	145	100
1176	gi17431512	Ralstonia solanacearum	PUTATIVE OUTER MEMBRANE CHANNEL LIPOPROTEIN TRANSMEMBRANE	71	26
1176	gi15823991	Streptomyces avermitilis	modular polyketide synthase	70	51
1177	AAM41939	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6870.	84	61
1177	gi870751	Homo sapiens	N-acetylgalactosamine 6-sulfate sulfatase (GALNS)	84	61
1177	gi618426	Homo sapiens	N-acetylgalactosamine 6-sulphatase	84	61
1178	gi435855	Mus sp.	CREB-binding protein; CBP	89	22
1178	AAW40058	Homo sapiens	USSH Cellular transcriptional factor CBP.	87	22
1178	gi17944308	Drosophila - melanogaster	RE12101p	86	26
1179	AAM25814	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1329.	73	93
1179	AAM25290	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:805.	73	93
1179	AAM79441	Homo sapiens	HYSE- Human protein SEQ ID NO 3087.	73	93
1180	AAB88388	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0131.	719	97
1180	gi20810493	Homo sapiens	Similar to RIKEN cDNA 2810417M05 gene	716	96
1180	AAD30543_aa1	Homo sapiens	MILL- Human B7RP-2 DNA.	83	38
1181	ABB14686	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 3343.	190	97
1181	gi14329731	Secale cereale	high molecular weight glutenin subunit x	88	27
1181	gi14329761	Triticum aestivum	high molecular weight glutenin subunit x	84	26
1182	gi11692645	Mus musculus	aspartyl beta-hydroxylase	74	28
1182	gi11878112	Mus musculus	aspartyl beta-hydroxylase 6.6 kb transcript	74	28
1182	gi11878110	Mus musculus	aspartyl beta-hydroxylase 4.5 kb transcript	74	28
1183	gi15485622	Homo sapiens	Q9H4T4 like	80	25
1183	gi19714949	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	TonB protein	78	32
1183	gi7717375	Homo sapiens	human CHD2-52 down syndrome cell adhesion molecule	71	23

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1184	AAU83667	Homo sapiens	GETH Human PRO protein, Seq ID No 152.	388	100
1184	AAG89161	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 281.	388	100
1184	AAU99348	Homo sapiens	GETH Human PRO1194 (UNQ607) amino acid sequence SEQ ID NO:29.	388	100
1185	AAB93506	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12830.	543	100
1185	AAB87570	Homo sapiens	GETH Human PRO1268.	426	95
1185	AAU78808	Homo sapiens	PROT- Hydrophobic domain containing protein clone HP10537 protein sequence.	426	95
1187	gi15823978	Streptomyces avermitilis	modular polyketide synthase	75	41
1187	AAB66657	Homo sapiens	HSCR- Human elastin protein without signal peptide.	71	39
1187	AAU69137	Homo sapiens	UNSY Amino acid sequence of a human tropoelastin derivative.	71	39
1188	gi6907090	Oryza sativa (japonica cultivar-group)	Similar to Oryza sativa root-specific RCc3 mRNA. (L27208)	76	30
1188	AAU36063	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 448.	74	26
1188	AAU35971	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 220.	73	26
1189	gi9827989	Leishmania major	possible CG12797 protein	72	36
1189	gi13625467 gb AAK35068.1	Leishmania donovani	LACK protective antigen	68	27
1190	gi17027071	Xiphocentron sp. UMSP00002937 2-Costa Rica	elongation factor-1 alpha	107	27
1190	gi310665	Strongylocentrotus purpuratus	Nf-Y-A subunit	88	24
1190	gi21743	Triticum aestivum	high molecular weight glutenin subunit 1Ax1	86	23
1191	gi16878287	Homo sapiens	Similar to C-terminal modulator protein	167	96
1191	gi15866714	Homo sapiens	C-terminal modulator protein	167	96
1191	AAO06984	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 20876.	132	83
1192	AAD05496_aal	Homo sapiens	HUMA- Human secreted protein-encoding gene 5 cDNA clone HHBCS39, SEQ ID NO:15.	859	100
1192	AAE01707	Homo sapiens	HUMA- Human gene 5 encoded secreted protein HHBCS39, SEQ ID NO:119.	859	100
1192	AAE01676	Homo sapiens	HUMA- Human gene 5 encoded secreted protein HHBCS39, SEQ ID NO:88.	859	100
1193	gi18650588	Homo sapiens	retinoic acid early transcript 1	1312	99
1193	AAB15540	Homo sapiens	INCY- Human immune system molecule from Incyte clone 3402252.	1283	97
1193	ABB84887	Homo sapiens	GETH Human PRO791 protein	1234	94

136  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			sequence SEQ ID NO:142.		
1195	gi1196427	Homo sapiens	gag 2 protein	248	50
1195	gi1780975	Human endogenous retrovirus K	gag protein	248	50
1195	gi1556397	Human endogenous retrovirus K	gag	248	50
1196	gi556256	Leishmania donovani	G protein alpha subunit	72	22
1197	AAAY07237	Homo sapiens	ISTF Wild type monocyte chemotactic protein 2.	121	100
1197	AAAY05300	Homo sapiens	ISTF C-C chemokine, MCP2.	121	100
1197	AAW42072	Homo sapiens	INCY- Human MC proprotein.	121	100
1198	ABB57423	Homo sapiens	HUMA- Human secreted protein encoding polypeptide SEQ ID NO 69.	187	79
1198	ABB57394	Homo sapiens	HUMA- Human secreted protein encoding polypeptide SEQ ID NO 40.	187	79
1198	AAAY59757	Homo sapiens	META- Human normal ovarian tissue derived protein 34.	187	79
1199	AAAY72603	Homo sapiens	INCY- Human Electron Transfer Protein, ETRN-1.	155	100
1199	AAB88465	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0259.	155	100
1199	AAE03926	Homo sapiens	HUMA- Human gene 29 encoded secreted protein HTADC63, SEQ ID NO:89.	155	100
1200	gi6458884	Deinococcus radiodurans	chorismate mutase/prephenate dehydratase	73	42
1201	gi20803920	Mesorhizobium loti	HYPOTHETICAL PROTEIN	68	32
1201	gi 17545158 ref NP_518560.1	Ralstonia solanacearum	PUTATIVE LIPASE/ESTERASE PROTEIN	66	31
1202	AAM67586	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27892.	69	30
1202	AAM55191	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 27296.	69	30
1202	gi849219	Saccharomyces cerevisiae	Pro1p: Glutamate 5-kinase (Swiss Prot. accession number P32264)	69	33
1203	gi18676554	Homo sapiens	FLJ00174 protein	269	84
1203	gi 20913341 ref XP_126763.1	Mus musculus	similar to FLJ00174 protein	125	81
1203	gi 20850247 ref XP_136664.1	Mus musculus	similar to proline-rich protein	121	33
1204	AAM68056	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 28362.	140	84
1204	AAM55676	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID	140	84

137

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO: 27781.		
1205	gi541624	Drosophila virilis	pdm2	71	39
1205	gi9955855	Aspergillus oryzae	RNA polymerase II largest subunit	69	38
1205	gi662296	Rattus norvegicus	MIBP1	68	32
1206	ABB50703	Homo sapiens	HUMA- Human secreted protein encoded by gene 52 SEQ ID NO:651.	260	94
1206	AAW88802	Homo sapiens	HUMA- Polypeptide fragment encoded by gene 52.	260	94
1206	ABB50706	Homo sapiens	HUMA- Human secreted protein encoded by gene 52 SEQ ID NO:654.	143	96
1207	AAM79588	Homo sapiens	HYSE- Human protein SEQ ID NO 3234.	72	41
1207	AAM78604	Homo sapiens	HYSE- Human protein SEQ ID NO 1266.	72	41
1207	AAB58944	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 652.	72	41
1208	AAE03429	Homo sapiens	HUMA- Human gene 3 encoded secreted protein HETDB76, SEQ ID NO: 112.	575	64
1208	gi19110438	Homo sapiens	polycystin-1L1	575	64
1208	AAE03463	Homo sapiens	HUMA- Human gene 3 encoded secreted protein HETDB76, SEQ ID NO: 146.	185	97
1209	gi6760015	Homo sapiens	brain protein	1114	85
1209	gi1747306	Mus musculus	SDR2	151	31
1209	gi20381292	Mus musculus	stromal cell derived factor receptor 2	151	31
1211	gi14043211	Homo sapiens	Similar to RIKEN cDNA 4931428F04 gene	460	89
1211	gi190508	Homo sapiens	salivary proline-rich protein precursor	113	28
1211	gi12862320	Homo sapiens	WDC146	102	28
1212	AAO14407	Homo sapiens	FARB Human 11 beta-hydroxysteroid dehydrogenase 1-like enzyme.	291	63
1212	AAM79592	Homo sapiens	HYSE- Human protein SEQ ID NO 3238.	217	45
1212	gi4581319	Homo sapiens	dJ28O10.3(HSD11B1 (hydroxysteroid (11-beta) dehydrogenase 1)	217	45
1213	AAR06514	Homo sapiens	STRI Natural human Platelet Factor-4var1 encoded by EcoRi fragment.	238	64
1213	gi292390	Homo sapiens	platelet factor 4	238	64
1213	AAZ28361_aal	Homo sapiens	SMIK Platelet factor-4 (PF-4) nucleotide sequence.	200	56
1214	AAD12580_aal	Homo sapiens	SAGA Human protein having hydrophobic domain encoding cDNA clone HP10753.	162	82
1214	AAD08193_aal	Homo sapiens	HUMA- Human secreted protein-encoding gene 3 cDNA clone HNTAC64, SEQ ID NO:13.	162	82
1214	AAD05544_aal	Homo sapiens	HUMA- Human secreted protein-encoding gene 12 cDNA clone HNTAC64, SEQ ID NO:63.	162	82



Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1215	gi21429094	Drosophila melanogaster	LD38004p	354	49
1215	gi15292155	Drosophila melanogaster	LD40717p	354	49
1215	AAG75596	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6360.	294	50
1216	gi7248894	Xenopus laevis	Arg protein-tyrosine kinase	84	35
1216	gi402191	Mus musculus	HNF-3beta	80	26
1216	gi404764	Mus musculus	fork head related protein	80	26
1218	AAM39205	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2350.	559	74
1218	AAO03505	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 17397.	502	81
1218	AAM40991	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5922.	467	66
1220	AAO01188	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15080.	248	86
1220	AAV73334	Homo sapiens	INCY- HTRM clone 1805061 protein sequence.	79	35
1220	gi20249	Oryza sativa	gt-2	77	32
1221	gi4519619	Haliothis discus	collagen pro alpha-chain	90	28
1221	gi7380690	Neisseria meningitidis Z2491	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentape pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	90	37
1221	gi7225645	Neisseria meningitidis MC58	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	90	37
1222	ABA05334_aa1	Homo sapiens	MILL- Human fucosyltransferase family member 32132 coding sequence.	2154	99
1222	AAM47905	Homo sapiens	MILL- Human fucosyltransferase family member 32132.	2154	99
1222	ABA05333_aa1	Homo sapiens	MILL- Human fucosyltransferase family member 32132 encoding cDNA.	2154	99
1223	AAV21852	Homo sapiens	INCY- Human signal peptide-containing protein (SIGP) (clone ID 2652271).	150	100
1223	AAV48563	Homo sapiens	META- Human breast tumour-associated protein 24.	150	100
1223	AAW75103	Homo sapiens	HUMA- Human secreted protein encoded by gene 47 clone HMCBP63.	150	100
1224	AAM67078	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27384.	517	99
1224	AAM54676	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26781.	517	99
1224	gi17467358	Sus scrofa	MIF2 suppressor	184	80
1225	gi9454237	Cochliobolus sativus	DNA binding protein MAT-1	73	30
1225	gi21428792	Drosophila melanogaster	GH03582p	72	38



140

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			656.		
1235	AAU18012	Homo sapiens	HUMA- Human immunoglobulin polypeptide SEQ ID No 157.	178	83
1235	ABB89226	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1602.	78	82
1236	gi10566951	Rattus norvegicus	s-gicerin/MUC18	85	45
1236	gi10566949	Rattus norvegicus	l-gicerin/MUC18	85	45
1236	AAB90798	Homo sapiens	NOJ/ Human shear stress-response protein SEQ ID NO: 96.	84	42
1238	gi21464300	Drosophila melanogaster	GH20068p	95	36
1238	gi3868879	Xenopus laevis	Zic-related-2	88	35
1238	gi1841756	Mus musculus	GATA-5 cardiac transcription factor	87	52
1239	gi17946266	Drosophila melanogaster	RE61793p	96	40
1239	gi15636898	Gallus gallus	formin binding protein 11-related protein	91	27
1239	gi780454	African swine fever virus	pB407L	88	30
1240	AAE05302	Homo sapiens	MILL- Human TANGO 457 protein.	1331	100
1240	AAE05303	Homo sapiens	MILL- Human mature TANGO 457 protein.	1207	100
1240	AAE05305	Homo sapiens	MILL- Human TANGO 457 protein cytoplasmic domain.	1201	100
1241	gi5640111	Lycopersicon esculentum	RAD23 protein	84	25
1241	gi17131739	Nostoc sp. PCC 7120	polyketide synthase type I	76	33
1241	gi 5640111 emb CAB51544.1	Lycopersicon esculentum	RAD23 protein	84	25
1242	AAG03496	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7577.	67	39
1242	gi 13876270 gb AAK26055.1	Mus musculus	protocadherin alpha 8	66	35
1243	AAE16665	Homo sapiens	MILL- Human calcium channel family member, 21784 protein.	196	87
1243	AAB62248	Homo sapiens	WARN Human calcium channel alpha2delta subunit.	196	87
1243	AAAY92320	Homo sapiens	WARN Human alpha-2-delta-C calcium channel subunit polypeptide.	196	87
1244	gi 4102990 gb AAD01637.1	Aspergillus nidulans	DNA polymerase epsilon homolog	70	30
1245	gi5917666	Zea mays	extensin-like protein	94	26
1245	gi19481644	shrimp white spot syndrome virus	WSSV052	89	36
1245	gi17016928	shrimp white spot syndrome virus	wsv001	89	36

141

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1246	AAO12623	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26515.	169	69
1246	AAO12822	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26714.	153	75
1246	AAO02255	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 16147.	123	65
1247	gi1653353	Synechocystis sp. PCC 6803	nodulation protein	75	28
1247	gi4468626	Mus musculus	TEF-5	74	26
1247	gi17430764	Ralstonia solanacearum	SKWP PROTEIN 5	74	23
1248	gi15139973	Sinorhizobium meliloti	CONSERVED HYPOTHETICAL PROTEIN	77	47
1249	gi7191078	Leishmania major	L712.2	99	29
1249	gi17384256	Homo sapiens	mucin 5	85	31
1249	gi5821153	Homo sapiens	RNA binding protein	83	33
1250	AAY36495	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 27.	124	86
1250	AAO12122	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26014.	123	91
1250	AAB95063	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16901.	121	90
1252	gi 15839838 ref NP_334875.1	Mycobacterium tuberculosis CDC1551	membrane protein, MmpL family	68	27
1254	AAG00399	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4480.	328	100
1254	gi21428466	Drosophila melanogaster	LD22609p	85	24
1254	gi19914274	Methanosarcina acetivorans str. C2A	sensory transduction histidine kinase [Methanosarcina	85	26
1256	gi14161094	Choloepus didactylus	von Willebrand Factor	80	24
1256	gi14161092	Cyclopes didactylus	von Willebrand Factor	78	23
1256	gi13872552	Acomys cahirinus	von Willebrand Factor	77	23
1258	gi7008025	Callithrix jacchus	prochymosin	715	64
1258	gi11990126	Camelus dromedarius	chymosin	634	57
1258	gi491952	synthetic construct	preprochymosin	618	56
1259	gi 21402709 ref NP_658694.1	Bacillus anthracis A2012	AMP-binding, AMP-binding enzyme [Bacillus anthracis	72	34
1260	gi 4505431 ref NP_002510.1	Homo sapiens	nuclear protein, ataxia-telangiectasia locus; NPAT gene; E14 gene	64	33
1260	gi 15309894 ref XP_040846.2	Homo sapiens	similar to nuclear protein, ataxia-telangiectasia locus; NPAT gene; E14 gene	64	33

142  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1260	gi 1304114 d bj BAA1186 1.1	Homo sapiens	NPAT	64	33
1261	gi4519535	Homo sapiens	Leukotriene B4 omega-hydroxylase	133	49
1261	gi1857022	Homo sapiens	leukotriene B4 omega-hydroxylase	133	49
1261	gi18266446	Homo sapiens	cytochrome P450, subfamily IVF, polypeptide 2	133	49
1262	gi13363530	Escherichia coli O157:H7	cell division protein HflB/FtsH protease	79	26
1262	gi746401	Escherichia coli	ATP-binding protein	79	26
1262	gi146028	Escherichia coli	ftsH	79	26
1263	AAW67859	Homo sapiens	HUMA- Human secreted protein encoded by gene 53 clone HBMCL41.	283	100
1264	gi11066248	Helix lucorum	presenilin	85	21
1264	gi 19115422  ref NP_5945 10.1	Schizosaccharom yces pombe	ribonuclease II RNB family protein; dis3-like	69	30
1264	gi 14720912  ref XP_0382 04.1	Homo sapiens	similar to Matrin 3	69	32
1265	gi5757703	Mus musculus	syntrophin-associated serine-threonine protein kinase	82	38
1265	gi4996035	Human herpesvirus 6	69.8% identical to U47 gene of strain U1102 of HHV-6	76	42
1265	gi330951	Gallid herpesvirus 1	ICP4	76	36
1266	gi 17511177  ref NP_4933 24.1	Caenorhabditis elegans	ZK1053.3.p	75	40
1266	gi 17538077  ref NP_4951 59.1	Caenorhabditis elegans	ZK1248.2.p	69	34
1267	gi915540	Ovis aries	pregnancy-specific antigen	85	25
1267	gi6179989	Capra hircus	pregnancy-associated glycoprotein-2	84	25
1267	gi9798658	Rhinolophus ferrumequinum	pepsinogen A	80	23
1268	gi 15789526  ref NP_2793 50.1	Halobacterium sp. NRC-I	serine proteinase; HtrA	69	30
1269	gi9988674	Influenza A virus (A/Swine/Wisco nsin/14094/99(H 3N2))	hemagglutinin protein	70	24
1269	gi6552676	Influenza A virus (A/Bangkok/1/97 (H3N2))	hemagglutinin	70	25
1269	gi6552638	Influenza A virus (A/Trinidad/51/9 6(H3N2))	hemagglutinin	70	24
1270	gi3378527	Zea mays	anther specific protein	87	41
1270	AAW15787	Homo sapiens	PENN- Human metastasis suppressor KiSS-1.	85	28
1270	gi21410770	Homo sapiens	Similar to RIKEN cDNA 1500005K14 gene	84	46

143

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1271	gi1335527	Human poliovirus 1	reading frame VP3	75	38
1271	gi61253	Human poliovirus 1	polyprotein	75	38
1271	gi17453412 ref XP_063132.1	Homo sapiens	similar to 60S ribosomal protein L7A (Surfeit locus protein 3)	76	40
1272	AAU87081	Homo sapiens	BRIM Sialic acid-binding Ig-related lectin, Siglec-11.	69	43
1272	AAU87077	Homo sapiens	BRIM Sialic acid-binding Ig-related lectin, Siglec-BMS-L3d.	69	43
1272	AAU87076	Homo sapiens	BRIM Sialic acid-binding Ig-related lectin, Siglec-BMS-L3c.	69	43
1273	AAA09121_aa1	Homo sapiens	CURA- Clone 2355875 cDNA (update), encodes syncollin homologue.	720	100
1273	AAAY92233	Homo sapiens	CURA- Clone 2355875f - syncollin homologue.	720	100
1273	AAB54267	Homo sapiens	HUMA- Human pancreatic cancer antigen protein sequence SEQ ID NO:719.	715	100
1274	gi15559064	Mus musculus	SNAG1	198	59
1274	AAU17435	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 1000.	131	62
1274	AAW99023	Homo sapiens	MOUN 17G2 peptide sequence.	131	62
1275	gi6753732 ref NP_034243.1	Mus musculus	epidermal growth factor	65	30
1275	gi50801 emb CAA24115.1	Mus musculus	polyprotein	65	30
1275	gi20341089 ref XP_109385.1	Mus musculus	epidermal growth factor	65	30
1276	AAM39205	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2350.	447	78
1276	AAM40991	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5922.	424	74
1276	AAO07159	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 21051.	401	75
1277	gi13905120	Mus musculus	RIKEN cDNA 0610013117 gene	134	35
1277	gi13936283	Mus musculus	TRH3	134	35
1277	AAB92625	Homo sapiens	HELI- Human protein sequence SEQ ID NO:10921.	127	35
1279	AAM66940	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27246.	362	85
1279	AAM54534	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26639.	362	85
1279	gi208153 gb AAA73184.1	synthetic construct	crystal toxin	79	40
1280	AAE05187	Homo sapiens	INCY- Human drug metabolising enzyme (DME-18) protein.	484	100

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1280	AAU12266	Homo sapiens	GETH Human PRO5780 polypeptide sequence.	484	100
1280	AAV91631	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 24 SEQ ID NO:304.	484	100
1281	AAH46856_aa1	Homo sapiens	HUMA- Human serine/threonine phosphatase encoding cDNA (clone ID HLDOO20).	238	100
1281	AAG77801	Homo sapiens	HUMA- Human HLDOO20 serine/threonine phosphatase protein sequence.	238	100
1281	AAB85476	Homo sapiens	HUMA- Human serine/threonine phosphatase (clone ID HLDOO20).	238	100
1282	gi 14762786 ref XP_047871.1	Homo sapiens	GS2 gene	70	30
1283	gi3860165	Arabidopsis thaliana	disease resistance protein RPP1-WsB	69	38
1283	AAO09033	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 22925.	68	38
1283	gi6967115	Arabidopsis thaliana	disease resistance protein homolog	68	38
1285	gi1055252	Rattus norvegicus	pheromone receptor VN5	78	32
1285	gi2746733	Drosophila virilis	circadian clock protein	73	26
1285	gi2641617	Drosophila virilis	TIM	73	26
1286	gi6013135	Rattus norvegicus	coxsackie-adenovirus-receptor homolog	86	67
1286	AAV50429_aa1	Homo sapiens	UYNV Human coxsackievirus and Ad2 and Ad5 receptor (HCAR) cDNA.	83	75
1286	AAV28845_aa1	Homo sapiens	DAND Human coxsackievirus and adenovirus receptor encoding DNA.	83	75
1287	AAU83224	Homo sapiens	ZYMO Novel secreted protein Z930757G12P.	642	100
1287	AAV70692	Homo sapiens	DAND Human soluble attractin-2.	84	54
1287	AAV70691	Homo sapiens	DAND Human membrane attractin-2.	84	54
1288	AAW70326	Homo sapiens	GEMY Secreted protein DU123 1.	1655	99
1288	ABB12473	Homo sapiens	HYSE- Human bone marrow expressed protein SEQ ID NO: 312.	547	72
1288	gi5689736	Homo sapiens	Myopodin protein	475	100
1289	gi4103543	Tomato chlorosis virus	heat shock protein 70	73	29
1289	gi12247413	Cristatella mucedo	cytochrome b	72	30
1289	gi 4103543 gb AAD01790.1	Tomato chlorosis virus	heat shock protein 70	73	29
1291	AAB94128	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14383.	520	98
1291	AAV85576	Homo sapiens	JANC Hs-UNC-53/1 fragment/GFP fusion insert of plasmid pGI3150.	520	98
1291	AAV85564	Homo sapiens	JANC Human homologue of UNC-53	520	98

145

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			(Hs-UNC-53/1) sequence.		
1292	AAAY01413	Homo sapiens	HUMA- Secreted protein encoded by gene 31 clone HHBAG64.	207	97
1292	AAAY05324	Homo sapiens	GEMY Human secreted protein ij167 5.	207	97
1292	gi15157864	Agrobacterium tumefaciens str. C58 (Cereon)	AGR_C_4816p	71	34
1294	AAB12146	Homo sapiens	PROT- Hydrophobic domain protein from clone HP10672 isolated from Thymus cells.	219	100
1295	gi17228767 ref NP_485315.1	Nostoc sp. PCC 7120	probable glycogen phosphorylase	78	34
1295	gi10835203 ref NP_001127.1	Homo sapiens	advanced glycosylation end product-specific receptor	65	58
1295	gi190846 gb AAA03574.1	Homo sapiens	receptor for advanced glycosylation end products	65	58
1296	gi17511816	Homo sapiens	Similar to RIKEN cDNA 1110032O22 gene	1268	99
1296	AAB88440	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0222.	688	100
1296	gi7211438	Homo sapiens	golgin-67	94	30
1298	gi18314436	Homo sapiens	Similar to RIKEN cDNA 4921511C04 gene	481	79
1298	gi1872546	Mus musculus	NIK	86	25
1298	gi5533305	Homo sapiens	somatostatin receptor interacting protein splice variant a	85	29
1299	gi1334643	Xenopus laevis	APEG precursor protein	105	27
1299	gi17428053	Ralstonia solanacearum	PROBABLE RIBONUCLEASE E (RNASE E) PROTEIN	100	32
1299	gi6690017	Herpesvirus papio	NTR	96	25
1300	AAB87346	Homo sapiens	HUMA- Human gene 5 encoded secreted protein HDPIE85, SEQ ID NO:87.	586	74
1300	AAB44298	Homo sapiens	GETH Human PRO706 (UNQ370) protein sequence SEQ ID NO:385.	586	74
1300	AAAY41742	Homo sapiens	GETH Human PRO706 protein sequence.	586	74
1301	gi218572	Pan troglodytes	prot GOR	1344	62
1301	gi243898	Pan	GOR	1040	68
1301	gi17862570	Drosophila melanogaster	LD38414p	486	45
1302	gi13276598	Homo sapiens	dJ614O4.7 (Novel protein)	260	28
1302	gi13397804	Homo sapiens	dJ616B8.3 (novel gene)	230	30
1302	AAB56641	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1219.	226	30
1303	gi603989	Drosophila melanogaster	salivary gland glue protein	149	23
1303	gi13324584	Borrelia burgdorferi	LMP1	129	17



146

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1303	gi161956	Trypanosoma cruzi	surface antigen	128	13
1304	gi13569248	Human immunodeficiency virus type 1	gag protein	81	34
1304	gi4324832	Human immunodeficiency virus type 1	gag-pol polyprotein	80	29
1304	gi11691875	Mus musculus	ADP-ribosylation factor 1 GTPase activating protein	79	22
1305	AAO06469	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 20361.	191	100
1305	gi3608368	Xenopus laevis	origin recognition complex associated protein p81	69	30
1305	ABB15196	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 3853.	68	36
1306	AAE03657	Homo sapiens	INCY- Human extracellular matrix and cell adhesion molecule-21 (XMAD-21).	109	27
1306	ABB11890	Homo sapiens	HYSE- Human protocadherin Flamingo 1 homologue, SEQ ID NO:2260.	109	27
1306	gi3449298	Homo sapiens	MEGF2	109	27
1308	gi9294050	Arabidopsis thaliana	protein kinase-like protein	84	32
1308	gi15983765	Arabidopsis thaliana	AT3g24550/MOB24_8	84	32
1308	gi13877617	Arabidopsis thaliana	protein kinase-like protein	84	32
1309	AAU00375	Homo sapiens	BERN/ Human stem cell growth factor receptor.	127	54
1309	AAE07145	Homo sapiens	SALK Human Kit/stem cell factor receptor kinase insert region.	127	54
1309	gi3236223	Equus caballus	tyrosine kinase receptor homolog	127	50
1310	gi21449343	Actinosynnema pretiosum subsp. auranticum	polyketide synthase	77	46
1310	gi21114513	Xanthomonas campestris pv. campestris str. ATCC 33913	transcriptional regulator	75	36
1310	gi13364364	Escherichia coli O157:H7	acetylglutamate kinase	73	36
1311	gi20146220	Oryza sativa (japonica cultivar-group)	similar to splicing factor/activator protein	110	33
1311	gi206712	Rattus norvegicus	salivary proline-rich protein	104	27
1311	AAV84592	Homo sapiens	UNIW Amino acid sequence of a human artemin polypeptide.	103	34
1312	gi2065210	Mus musculus	Pro-Pol-dUTPase polyprotein	530	69
1312	gi10834720 gb AAG23790.1 AF258	Homo sapiens	PP565	249	66

147

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	587 1				
1312	gi 13194728 gb AAK15526.1 AF329451.1	Gallus gallus	pol-like protein ENS-3	115	21
1313	AAW03515	Homo sapiens	SHKJ Human DOCK180 protein.	147	58
1313	gi1339910	Homo sapiens	DOCK180 protein	147	58
1313	gi1504002	Homo sapiens	similar to a human major CRK-binding protein DOCK180.	111	43
1314	gi12007418	Mus musculus	B3 olfactory receptor	76	38
1314	gi18480290	Mus musculus	olfactory receptor MOR260-3	76	38
1314	gi12007432	Mus musculus	B3 olfactory receptor	76	38
1315	gi483581	Mus musculus	Notch 3	82	26
1315	gi18159668	Pyrobaculum aerophilum	paREP2b	81	29
1315	gi4584086	Spermatozopsis similis	p210 protein	79	25
1316	AAM71305	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 31611.	422	98
1316	AAM58790	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 30895.	422	98
1316	gi149490	Lactococcus lactis	sucrose-6-phosphate hydrolase	72	31
1317	gi1620040	Paramecium bursaria Chlorella virus 1	Asp-rich	72	28
1317	gi3721615	Cyprinus carpio	MEF2C	71	25
1317	gi 9631936 ref NP_048725.1	Paramecium bursaria Chlorella virus 1	Asp-rich	72	28
1318	gi 21291797 gb EAA03942.1	Anopheles gambiae str. PEST	agCP3974	74	35
1319	gi21306283	Chlamydomonas reinhardtii	iron transporter Ftr1	74	30
1319	AAB60461	Homo sapiens	INCY- Human cell cycle and proliferation protein CCYPR-9, SEQ ID NO:9.	73	33
1319	gi6013155	Homo sapiens	p35srj	73	33
1320	gi9717245	Mus musculus	cytoplasmic dynein heavy chain	430	94
1320	gi402528	Rattus norvegicus	cytoplasmic dynein heavy chain	430	94
1320	gi294543	Rattus norvegicus	dynein heavy chain	430	94
1323	gi 17221411 emb CAD12639.1	Burkholderia cepacia	kdo transferase	70	34
1324	gi1698601	Cricetulus griseus	beta-1,6-N-acetylglucosaminyltransferase	440	38
1324	gi349091	Rattus norvegicus	N-acetylglucosaminyltransferase V	438	43
1324	gi18997007	Mus musculus	N-acetylglucosaminyltransferase V	438	43

148

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1325	AAM70545	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 30851.	115	47
1325	AAM58098	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 30203.	115	47
1325	AAM72994	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 33300.	111	28
1326	gi12724969	Lactococcus lactis subsp. lactis	phenolic acid decarboxylase	77	46
1327	AAB53097	Homo sapiens	GETH Human angiogenesis-associated protein PRO1246, SEQ ID NO:167.	372	63
1327	AAU12416	Homo sapiens	GETH Human PRO1246 polypeptide sequence.	372	63
1327	AAV99377	Homo sapiens	GETH Human PRO1246 (UNQ630) amino acid sequence SEQ ID NO:132.	372	63
1328	gi6014505	Hepatitis GB virus B	polyprotein	76	43
1328	gi765145	Hepatitis GB virus B	polypeptide	68	41
1328	gi 20544059 ref XP_086220.4	Homo sapiens	similar to U4/U6-associated RNA splicing factor	294	100
1329	AAV42689_aa1	Homo sapiens	SIBI- DNA encoding human calcium channel alpha-2 subunit.	158	91
1329	AAQ84667_aa1	Homo sapiens	SALK Human neuronal calcium channel subunit alpha 2c.	158	91
1329	AAQ84664_aa1	Homo sapiens	SALK Human neuronal calcium channel subunit alpha 2b.	158	91
1330	gi19923	Nicotiana tabacum	pistil extensin like protein, partial CDS	71	38
1330	gi 144429 gb AAA56792.1	Cellulomonas fimi	beta-1,4-xylanase	67	30
1331	gi2388676	Mytilus edulis	precollagen P	85	35
1331	gi17862044	Drosophila melanogaster	LD06016p	75	30
1331	gi13879780	Mycobacterium tuberculosis CDC1551	PE_PGRS family protein	74	30
1333	AAO00015	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 13907.	442	61
1333	AAB82479	Homo sapiens	ZYMO Human RING finger protein Zapop2.	81	31
1333	gi20975274	Homo sapiens	skeletrophin	81	31
1334	ABB11819	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2189.	367	82
1334	AAW80398	Homo sapiens	GEMY A secreted protein encoded by clone cw1543_3.	130	67
1334	gi5081693	Samanea saman	pulvinus inward-rectifying channel SPICK2	70	34
1335	ABB89969	Homo sapiens	HUMA- Human polypeptide SEQ ID	142	96

149

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO 2345.		
1335	AAB38385	Homo sapiens	HUMA- Human secreted protein encoded by gene 18 clone HTLEJ24.	142	96
1335	AAB38338	Homo sapiens	HUMA- Human secreted protein encoded by gene 18 clone HTLFE57.	142	96
1336	gi 14590195 ref NP_142260.1	Pyrococcus horikoshii	asparaginyl-tRNA synthetase	70	37
1337	gi3879419	Caenorhabditis elegans	contains similarity to Pfam domain: PF00102 (Protein-tyrosine phosphatase), Score=51.6, E-value=1.8e-14, N=1	69	29
1337	gi 17563828 ref NP_505965.1	Caenorhabditis elegans	protein tyrosine phosphatase	69	29
1338	gi 2072960 gb AAC51268.1	Homo sapiens	p40	138	33
1338	gi 4185940 emb CAA76880.1	Human endogenous retrovirus K	env protein	124	75
1338	gi 757872 emb CAA57723.1	Human endogenous retrovirus	env	124	75
1340	gi1491979	Molluscum contagiosum virus subtype 1	MC036R	78	33
1340	gi 9628968 ref NP_043987.1	Molluscum contagiosum virus	MC036R	78	33
1341	gi18676514	Homo sapiens	FLJ00154 protein	1560	100
1341	AAB84252	Homo sapiens	HUMA- Amino acid sequence of a human cytokine receptor-like protein.	572	63
1341	AAB84251	Homo sapiens	HUMA- Human cytokine receptor-like protein fragment.	572	63
1342	AAY27757	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 47.	152	71
1342	AAB27551	Homo sapiens	MYRI- Human tumour suppressor BRG1 encoded by cDNA mutated at base 1705.	77	32
1342	AAB27550	Homo sapiens	MYRI- Human tumour suppressor BRG1 protein from cell lines DU145 and NCI-H1300.	77	32
1344	gi21464394	Drosophila melanogaster	RE18651p	78	26
1344	AAM39065	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2210.	77	21
1344	gi338290	Homo sapiens	son3 protein	77	21
1345	gi2202	Canis sp.	Clox	135	37
1345	gi3879551	Caenorhabditis elegans	contains similarity to Pfam domain: PF01391 (Collagen triple helix repeat (20 copies)), Score=56.4, E-value=2e-13, N=2; PF01484 (Nematode cuticle collagen N-terminal domain),	125	33

150

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			Score=87.2, E-value=1.1e-22, N=1		
1345	gi158695	Drosophila melanogaster	tropomyosin isoform 33 (9C)	118	30
1346	gi7862077	Giardia intestinalis	3-hydroxy-3-methylglutaryl-coenzyme A reductase	90	26
1346	gi1098615	Mycoplasma pneumoniae	adhesin-related 30 kDa protein	87	23
1346	gi20380058	Homo sapiens	Similar to PRAM-1 protein	84	28
1347	gi13905302	Mus musculus	Similar to ATPase, class II, type 9A	736	85
1347	gi17862322	Drosophila melanogaster	LD22119p	633	72
1347	AAM25271	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:786.	572	100
1348	gi456319	Bacteriophage FC1	74kDa protein	75	33
1348	gi1524115	Lycopersicon esculentum	subtilisin-like endoprotease	73	28
1348	gi4200334	Lycopersicon esculentum	P69A protein	73	28
1349	gi21391988	Drosophila melanogaster	HL08052p	78	31
1349	gi20148339	Arabidopsis thaliana	cyclin delta-3	77	25
1349	gi 17647607 ref NP_523423.1	Drosophila melanogaster	maroon-like; bronzy; section 5	78	31
1351	gi18676524	Homo sapiens	FLJ00159 protein	164	52
1351	gi21392066	Drosophila melanogaster	RE04357p	139	34
1351	AAB92637	Homo sapiens	HELI- Human protein sequence SEQ ID NO:10953.	81	43
1352	gi19071965	Aspergillus oryzae	chitin synthase	79	28
1352	gi17945592	Drosophila melanogaster	RE26660p	78	41
1352	gi16184663	Drosophila melanogaster	LD28370p	74	22
1353	gi 11037117 gb AAG27485.1 AF194537.1	Homo sapiens	NAG13	307	65
1353	gi 1335205 emb CAA36480.1	Homo sapiens	ORFII	305	65
1354	gi1388166	Drosophila melanogaster	Bowel	80	32
1354	gi15553187	Scyliorhinus canicula	homeodomain protein Otx1	79	22
1354	AAV85573	Homo sapiens	JANC Hs-UNC-53/3 fragment/GFP fusion insert of plasmid pGI3303.	78	26
1358	gi 21288288 gb EAA00609.1	Anopheles gambiae str. PEST	agCP9766	71	30
1358	gi 17465558	Homo sapiens	similar to mucin	68	36

151

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	ref XP_069888.1				
1359	gi 21302892 gb EAA15037.1	Anopheles gambiae str. PEST	agCP5020	70	31
1361	gi15080686	Lentinula edodes	CDC5	79	26
1361	gi495516	Plasmodium vivax	circumsporozoite protein	77	31
1361	gi21070569	Dictyostelium discoideum	VSAE2 (FRAGMENT). 3/101	76	31
1362	gi8953400	Arabidopsis thaliana	1-D-deoxyxylulose 5-phosphate synthase-like protein	73	23
1362	gi 15239030 ref NP_196699.1	Arabidopsis thaliana	1-D-deoxyxylulose 5-phosphate synthase - like protein	73	23
1363	gi2444430	Xenopus laevis	deacetylase	327	81
1363	gi602098	Xenopus laevis	yeast RPD3 homologue	324	80
1363	AAB49954	Homo sapiens	METH- Human histone deacetylase HDAC-1.	323	80
1364	AAM69686	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29992.	418	55
1364	AAM57281	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 29386.	418	55
1364	gi 1780971 emb CAA71416.1	Human endogenous retrovirus K	gag protein	172	37
1365	gi437084	Gallus gallus	vitamin D3 hydroxylase associated protein	510	41
1365	gi2149156	Homo sapiens	fatty acid amide hydrolase	477	38
1365	AAW57783	Homo sapiens	SCRI Human fatty acid amide hydrolase.	468	38
1366	gi3510695	Homo sapiens	DNA polymerase theta	77	21
1366	gi309132	Mus musculus	calnexin	72	22
1366	gi15214567	Mus musculus	Similar to calnexin	72	22
1367	gi 17508849 ref NP_491426.1	Caenorhabditis elegans	helicase	73	40
1368	gi5457567	Pyrococcus abyssi	Na <sup>+</sup> /H <sup>+</sup> antiporter (napA-1)	76	33
1368	gi8247211	Candida albicans	She9 protein	69	31
1368	gi 14590079 ref NP_142143.1	Pyrococcus horikoshii	Na <sup>(+)</sup> /H <sup>(+)</sup> antiporter	76	30
1369	gi17644260	Homo sapiens	bb206l21.1 (ATPase, Class VI, type 11C)	305	98
1369	AAO14200	Homo sapiens	INCY- Human transporter and ion channel TRICH-17.	166	50
1369	gi5080816	Arabidopsis thaliana	Putative ATPase	166	49
1370	gi 18573281 ref XP_095933.1	Homo sapiens	similar to 40S ribosomal protein S3A	70	38

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1372	gi6683562	Mus musculus	heparan sulfate 6-sulfotransferase 3	886	91
1372	gi6683558	Mus musculus	heparan sulfate 6-sulfotransferase 2	265	72
1372	ABL39900_aa1	Homo sapiens	SEGK Human HS6ST2v encoding cDNA SEQ ID NO:1.	262	71
1373	gi20882231 ref XP_139203.1	Mus musculus	similar to LIM domain only 7	76	24
1373	gi20302988 gb AAM18948.1 AF498989.1	Medicago sativa	nodule-specific glycine-rich protein 3	72	26
1373	gi9965267 gb AAG10008.1	infectious hypodermal and hematopoietic necrosis virus	non-structural protein 2	72	24
1374	gi3355835	Rhizobium etli	RBSK	78	32
1374	gi7453560	Polyangium cellulosum	epoD	73	28
1374	gi1749684	Schizosaccharomyces pombe	similar to Saccharomyces cerevisiae porphobilinogen deaminase, SWISS-PROT Accession Number P28789	72	28
1375	gi16973455	Danio rerio	beta-3-galactosyltransferase	1050	63
1375	AAB24035	Homo sapiens	GETH Human PRO4397 protein sequence SEQ ID NO:42.	725	46
1375	AAB88404	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0159.	709	43
1376	gi7668	Drosophila melanogaster	bsg25D protein	73	33
1376	gi20177037	Drosophila melanogaster	LD21844p	73	33
1376	gi1353669	Caenorhabditis elegans	UNC-24	69	43
1379	AAS16182_aa1	Homo sapiens	GENA- Human apolipoprotein C1 (APOC1) DNA.	245	67
1379	AAU10534	Homo sapiens	GENA- Human apolipoprotein C1 (APOC1) polypeptide.	245	67
1379	AAS16825_aa1	Homo sapiens	GENA- Human apolipoprotein C1 (APOC1) DNA coding sequence.	245	67
1380	AAV36290	Homo sapiens	HUMA- Human secreted protein encoded by gene 67.	177	74
1380	gi16551305	Tatianyx annacites	DNA-directed RNA polymerase beta' subunit 2	71	38
1380	gi3411013	Candida albicans	protein mannosyltransferase 1	68	35
1381	AAM80132	Homo sapiens	HYSE- Human protein SEQ ID NO 3778.	173	66
1381	gi4731867	Dictyostelium discoideum	sterol glucosyltransferase	107	30
1381	AAB74726	Homo sapiens	INCY- Human membrane associated protein MEMAP-32.	89	41
1382	AAB62100	Homo sapiens	WIST- Human bridging integrator-2 (Bin2) protein.	78	27
1382	gi6527168	Homo sapiens	breast cancer associated protein BRAP1	78	27
1382	gi5852834	Homo sapiens	bridging integrator-2	78	27

153  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1383	gi7670050	Xenopus laevis	type I collagen alpha 1	92	27
1383	AAO01606	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15498.	85	29
1383	gi17738485	Agrobacterium tumefaciens str. C58 (U. Washington)	biopolymer transport protein	85	28
1384	gi20451261	Caenorhabditis elegans	C. elegans GCY-17 protein (corresponding sequence W03F11.2)	71	26
1384	gi2665714	Agrobacterium tumefaciens	moaC	71	29
1384	gi 20864452 ref XP_150076.1	Mus musculus	RIKEN cDNA 2410018E23	130	59
1385	AAAY94938	Homo sapiens	GEMY Human secreted protein clone ye78_1 protein sequence SEQ ID NO:82.	103	25
1385	gi12831176	Agelaius phoeniceus	gamma filamin protein	96	29
1385	AAU81998	Homo sapiens	INCY- Human secreted protein SECP24.	87	27
1386	gi10440468	Homo sapiens	FLJ00070 protein	102	41
1386	gi11136912	Danio rerio	RPTP-alpha protein	94	32
1386	gi20377083	Homo sapiens	p78	92	36
1387	AAM40810	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5741.	190	59
1387	AAM39024	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2169.	190	59
1387	gi15080474	Homo sapiens	Similar to RIKEN cDNA 1700023O11 gene	190	59
1388	gi12802591	Bovine herpesvirus 4	tegument protein	82	30
1388	gi950226	Saccharomyces cerevisiae	Trf4p	73	26
1388	gi 13095641 ref NP_076556.1	Bovine herpesvirus 4	tegument protein	82	30
1389	AAI67224_aa1	Homo sapiens	CORI- B511S cDNA sequence.	363	100
1389	AAF85500_aa1	Homo sapiens	EOSB- Nucleotide sequence of a human breast cancer protein designated BCH1.	363	100
1389	AAA54120_aa1	Homo sapiens	EOSB- Breast cancer protein BCH1 coding sequence.	363	100
1390	gi184653	Homo sapiens	IFN-alpha responsive transcription factor	74	30
1390	gi 2580453 gb AAB82336.1	Xenopus laevis	Xbap	68	47
1391	AAB88456	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0246.	85	52
1391	AAB62392	Homo sapiens	LEXI- Human LDL receptor family protein (LDLP).	85	52
1392	ABB12009	Homo sapiens	HYSE- Human RAMP1 homologue,	90	100



154

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			SEQ ID NO:2379.		
1392	gi3171910	Homo sapiens	RAMP1	90	100
1392	gi12653551	Homo sapiens	receptor (calcitonin) activity modifying protein 1	90	100
1394	gi4467343	Drosophila melanogaster	EG:140G11.1	70	27
1394	gi6018879	Drosophila melanogaster	BACN4L24.d	70	27
1394	gi157993	Drosophila melanogaster	developmental protein	70	27
1395	gi4928919	Arabidopsis thaliana	zinc finger protein 2	86	26
1395	gi2702272	Arabidopsis thaliana	expressed protein	86	26
1396	AAM25276	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:791.	729	93
1396	AAE14340	Homo sapiens	INCY- Human protease PRTS-5 protein.	528	33
1396	AAB47561	Homo sapiens	INCY- Protease PRTS-3.	528	33
1397	gi18369843	Infectious salmon anemia virus	P6	89	40
1397	gi4092530	Infectious salmon anemia virus	NS1 protein	87	39
1397	gi14009648	Infectious salmon anemia virus	NS1	87	39
1398	AAW63707	Homo sapiens	UYOR- Human hSK2 protein.	331	91
1398	gi1575663	Rattus norvegicus	calcium-activated potassium channel rSK2	331	91
1398	gi15082148	Homo sapiens	small-conductance calcium-activated potassium channel	331	91
1399	AAB01381	Homo sapiens	INCY- Neuron-associated protein.	1653	68
1399	gi18157547	Mus musculus	pecanex-like 3	1620	66
1399	gi6650377	Mus musculus	pecanex 1	1277	51
1400	gi 20887681 ref XP_140575.1	Mus musculus	similar to melastatin 1	468	91
1400	gi 3243075 gb AAC8000.0.1	Homo sapiens	melastatin 1	355	75
1400	gi 20552333 ref XP_007662.9	Homo sapiens	similar to melastatin 1	355	75
1401	AAU15955	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 908.	931	92
1401	gi3978441	Homo sapiens	PITSLRE protein kinase alpha SV9 isoform	95	24
1401	gi1517914	Homo sapiens	monocytic leukaemia zinc finger protein	91	28
1402	gi1289326	Mus musculus	ROR-alpha 1	84	25
1402	gi530878	Chlamydomonas eugametos	amino acid feature: N-glycosylation sites, aa 41 .. 43, 46 .. 48, 51 .. 53, 72 ..	79	32

155

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score.	% Identity
			74, 107 .. 109, 128 .. 130, 132 .. 134, 158 .. 160, 163 .. 165; amino acid feature: Rod protein domain, aa 169 .. 340; amino acid feature: globular protein domain, aa 32 .. 168		
1402	gi220763	Rattus norvegicus	HES-3 factor	79	52
1403	gi 20479430 ref XP_114955.1	Homo sapiens	similar to olfactory receptor MOR231-1	71	32
1403	gi 20480897 ref XP_115014.1	Homo sapiens	similar to olfactory receptor MOR234-3	71	32
1404	AAA88548_aa1	Homo sapiens	SMIK Human CASB616 cDNA.	89	100
1404	AAB19591	Homo sapiens	SMIK Human CASB616.	89	100
1404	gi1100110	Homo sapiens	protein-tyrosine kinase	89	100
1405	gi4206753	Oryctolagus cuniculus	homeodomain-containing protein	74	24
1405	gi13445253	Mus musculus	orphan Gpr37-like protein 1	72	33
1405	gi3080552	Mus musculus	Hoxa-9	71	50
1406	AAM50585	Homo sapiens	NISB Benign prostatic hyperplasia associated protein JT460914.	325	100
1406	gi18031947	Homo sapiens	SOCS box protein ASB-5	325	100
1406	AAU20593	Homo sapiens	HUMA- Human secreted protein, Seq ID No 585.	316	100
1407	AAU83222	Homo sapiens	ZYMO Novel secreted protein Z930005G2P.	895	97
1407	AAY02712	Homo sapiens	HUMA- Human secreted protein encoded by gene 63 clone HBJFV28.	91	56
1407	AAO00641	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 14533.	86	64
1408	ABB17944	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 6601.	81	53
1408	AAM77906	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 38212.	72	40
1408	AAM65199	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 37304.	72	40
1409	gi5230847	Vitreoscilla sp. C1	glutamine synthetase homolog	68	33
1409	gi8515736	Drosophila melanogaster	highwire	67	35
1409	gi3138797	Sulfolobus shibatae	Ssh7b	65	48
1410	AAW23309	Homo sapiens	EIII- Human Werner's syndrome WS-2 protein.	151	96
1410	gi1913785	Homo sapiens	Rep-8	151	96
1410	gi18089098	Homo sapiens	reproduction 8	151	96
1411	gi 21297468 gb EAA09613.1	Anopheles gambiae str. PEST	agCP15537	166	56
1411	gi 20983200	Mus musculus	RIKEN cDNA 1810030O07	73	24

156

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	ref XP_135812.1				
1412	gi532572	Hordeum vulgare	lipoygenase 1	82	28
1412	gi945419	Mus musculus	hepatoma derived growth factor (HDGF)	77	35
1412	gi17932895	stork hepatitis B virus	preC/core antigen	77	26
1413	gi2370143	Homo sapiens	immunoglobulin-like domain-containing 1	169	42
1413	gi2645890	Homo sapiens	IGSF1	169	42
1413	AAB40232	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 46 SEQ ID NO:142.	162	40
1414	gi21204314	Staphylococcus aureus subsp. aureus MW2	proline-tRNA ligase	78	32
1414	gi14247033	Staphylococcus aureus subsp. aureus Mu50	proline-tRNA ligase	78	32
1414	gi13701063	Staphylococcus aureus subsp. aureus N315	proline-tRNA ligase	78	32
1415	gi9948469	Pseudomonas aeruginosa	probable non-ribosomal peptide synthetase	78	31
1415	AAE19251	Homo sapiens	BIOI- SOS1 protein sequence from PS462.	75	23
1415	AAU84311	Homo sapiens	BAAK/ Protein ABCB2 differentially expressed in breast cancer tissue.	74	30
1416	gi18676710	Homo sapiens	FLJ00254 protein	623	75
1416	gi2065210	Mus musculus	Pro-Pol-dUTPase polypeptide	583	69
1416	gi 18676710 dbj BAB85007.1	Homo sapiens	FLJ00254 protein	623	75
1417	AAR85785	Homo sapiens	UYNH Human GRB-10.	77	32
1417	gi841210	Mus musculus	growth factor receptor binding protein Grb10	77	32
1417	AAM90963	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:18556.	74	32
1419	AAM79990	Homo sapiens	HYSE- Human protein SEQ ID NO 3636.	82	100
1419	AAM79006	Homo sapiens	HYSE- Human protein SEQ ID NO 1668.	82	100
1419	AAR28494	Homo sapiens	XIAM/ Sequence encoded by the CAMPATH-1 antigen cDNA.	82	100
1420	AAU01383	Homo sapiens	MILL- Human TANGO 499 form 2, variant 1 amino acid sequence.	828	73
1420	AAU01382	Homo sapiens	MILL- Human TANGO 499 form 2, variant 4 amino acid sequence.	828	73
1420	AAU01380	Homo sapiens	MILL- Human TANGO 499 form 2, amino acid sequence.	828	73
1421	gi19069609	Encephalitozoon cuniculi	PROTEASOME REGULATORY SUBUNIT YTA6 OF THE AAA	76	26

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			FAMILY OF ATPASES		
1422	AAM66177	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26483.	199	72
1422	AAM53791	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25896.	199	72
1422	AAM68472	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 28778.	176	81
1423	gi1800227	Oryza sativa	Bowman-Birk proteinase inhibitor	74	34
1423	gi10141005	San Miguel sea lion virus	non-structural polyprotein	74	26
1423	gi17490177 refXP_062300.1	Homo sapiens	similar to RING finger protein 18 (Testis-specific ring-finger protein)	76	28
1424	gi461336	Pyrenomonas salina	hsp70	75	29
1424	gi13880037	Mycobacterium tuberculosis CDC1551	membrane protein, MmpL family	75	24
1424	gi1449306	Mycobacterium tuberculosis H37Rv	mmpL2	75	24
1425	gi15600	Enterobacteria phage T7	gene 7.3, host range	79	30
1425	gi16198065	Drosophila melanogaster	LD28477p	77	30
1425	gi11870012	Drosophila melanogaster	xnp/atx-x DNA helicase	77	30
1426	gi16185397	Drosophila melanogaster	LD39815p	204	44
1426	gi2244793	Arabidopsis thaliana	disease resistance N like protein	86	30
1426	AAU84280	Homo sapiens	BGHM Human endometrial cancer related protein, HERC1.	77	26
1427	AAV36302	Homo sapiens	HUMA- Human secreted protein encoded by gene 79.	183	79
1427	AAB88359	Homo sapiens	HELL- Human membrane or secretory protein clone PSEC0087.	178	80
1427	AAM41635	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6566.	178	80
1428	AAU82008	Homo sapiens	INCY- Human secreted protein SECP34.	114	64
1428	AAB32391	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 21 SEQ ID NO:77.	114	64
1428	AAV08306	Homo sapiens	FIBR- Human collagen IX alpha-3 chain protein.	74	45
1429	gi2792523	Ralstonia solanacearum	alternative RNA sigma factor RpoS	69	30
1429	gi17428221	Ralstonia solanacearum	RNA POLYMERASE SIGMA S (SIGMA-38) FACTOR TRANSCRIPTION REGULATOR	69	33

158

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			PROTEIN		
1429	gi 5032313 ref NP_004014.1	Homo sapiens	dystrophin Dp140bc isoform; Dystrophin (muscular dystrophy, Duchenne and Becker types)	73	26
1433	gi9954445	Rattus norvegicus	TEMO	171	62
1433	gi14030260	maize rayado fino virus	polyprotein	79	32
1433	AAB95656	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18419.	77	36
1434	AAR04212	Homo sapiens	CALB- Human 32K alveolar surfactant protein.	391	43
1434	AAP60661	Homo sapiens	KUSH/ Genomic sequence of human alveolar surfactant protein (hASP)encoded by genomic DNA.	386	43
1434	AAB58135	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 473.	366	42
1435	gi17224904	Mus musculus	immunoglobulin superfamily member 9	180	48
1435	gi20988778	Homo sapiens	Similar to immunoglobulin superfamily, member 9	173	53
1435	gi14149050	Drosophila melanogaster	turtle protein, isoform 4	114	36
1436	gi1465855	Caenorhabditis elegans	C. elegans PQN-57 protein (corresponding sequence R09F10.7)	85	23
1436	gi1465856	Caenorhabditis elegans	C. elegans PQN-56 protein (corresponding sequence R09F10.2)	85	23
1436	gi17864717	Mus musculus	hornerin	83	26
1437	gi 21292574 gb EAA04719.1	Anopheles gambiae str. PEST	agCP3449	66	33
1438	ABB10160	Homo sapiens	HUMA- Human cDNA SEQ ID NO: 468.	166	62
1438	gi9657279	Vibrio cholerae	aspartokinase II/homoserine dehydrogenase, methionine-sensitive	71	28
1439	gi4582571	Gallus gallus	Hyperion protein, 419 kD isoform	75	24
1439	gi13165	Oenothera biennis	ATPase alpha-subunit (aa 1-511)	72	26
1439	gi903838	Oenothera berteriana	F-1-ATPase alpha subunit	72	26
1440	gi4558758	Homo sapiens	testis-specific chromodomain Y-like protein	233	62
1440	gi4558762	Mus musculus	testis-specific chromodomain Y-like protein	231	36
1440	gi3342716	Homo sapiens	testis-specific ChromoDomain Y isoform 1	195	36
1441	gi155627	Acanthamoeba castellanii	myosin I heavy chain	118	42
1441	gi13093370	Mycobacterium leprae	initiation factor IF-2	116	33
1441	AAY20289	Homo sapiens	UYRO- Human apolipoprotein E mutant protein fragment 5.	114	39
1442	gi2253707	Mus musculus	Daxx	84	36
1442	gi1934970	Plasmodium falciparum	AARP1 protein	79	65

159

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1442	gi4050098	Mus musculus	Fas-binding protein	78	34
1443	gi2425111	Dictyostelium discoideum	ZipA	90	26
1443	AAAY06119	Homo sapiens	HARD Human CIITA interacting protein 104 (CIP104).	88	26
1443	gi5420387	Leishmania major	proteophosphoglycan	86	21
1444	gi893355	Acinetobacter baumannii	L-2,4-diaminobutyrate decarboxylase	77	26
1445	ABB55744	Homo sapiens	FECH/ Human polypeptide SEQ ID NO 94.	135	47
1445	AAU39035	Homo sapiens	GEMY Human secreted protein nh328_5.	135	47
1445	AAAY28679	Homo sapiens	GEMY Human nh328_5 secreted protein.	135	47
1446	gi19744390	Homo sapiens	retinoic acid inducible in neuroblastoma cells RAINB1d	247	54
1446	gi19744388	Homo sapiens	retinoic acid inducible in neuroblastoma cells RAINB1	247	54
1446	AAAY85565	Homo sapiens	JANC Human homologue of UNC-53 (Hs-UNC-53/2) sequence.	240	52
1447	AAU19716	Homo sapiens	HUMA- Human novel extracellular matrix protein, Seq ID No 366.	71	31
1447	gi18025476	cercopithecine herpesvirus 15	BPLF1	71	38
1447	AAS14575_aal	Homo sapiens	MILL- Human cDNA encoding G protein-coupled receptor, GPCR, 52872.	69	62
1448	gi14027507	Mesorhizobium loti	salicylate hydroxylase	69	31
1449	AAG64798	Homo sapiens	SREH- Human peptide methionine sulfoxide reductase (hPMSR).	192	71
1449	AAB81893	Homo sapiens	SEQU- Human genomic database related protein SEQ ID NO: 38.	192	71
1449	AAM42046	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6977.	192	71
1450	gi18249657	Mus musculus	NC8	1063	80
1450	gi406748	Mus musculus	zinc finger protein	250	37
1450	AAB43498	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:943.	249	37
1451	ABB89331	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1707.	732	88
1451	gi13421927	Caulobacter crescentus CB15	MaoC family protein	273	42
1451	gi19338616	Methylobacterium extorquens	R-specific enoyl-CoA hydratase	261	44
1452	gi 20908171 ref XP_139715.1	Mus musculus	similar to NADPH oxidase 3; NADPH oxidase catalytic subunit-like 3	68	30
1452	gi 17533619 ref NP_495516.1	Caenorhabditis elegans	F32A5.8.p	67	42
1453	gi 15614051 ref NP_2423	Bacillus halodurans	sodium-dependent phosphate transporter	65	34

160  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	54.1				
1454	gi 17551878 ref NP_499090.1	Caenorhabditis elegans	TPR Domain	76	29
1455	AAM40727	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5658.	191	56
1455	AAM38941	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2086.	191	56
1455	gi19702127	Homo sapiens	P-Rex1 protein	191	56
1456	ABB05666	Homo sapiens	GEHU- Human nucleic acid management protein clone amy2 11n4.	496	91
1456	AAE03372	Homo sapiens	HUMA- Human gene 18 encoded secreted protein fragment, SEQ ID NO:152.	496	91
1456	AAE03371	Homo sapiens	HUMA- Human gene 18 encoded secreted protein fragment, SEQ ID NO:150.	496	91
1457	AAM66940	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27246.	290	77
1457	AAM54534	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26639.	290	77
1457	AAM64410	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 36515.	287	77
1458	AAB53445	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:985.	335	100
1458	AAY30055	Homo sapiens	ARIA- Amino acid sequence of a FK506-binding protein (FKBP).	165	91
1458	AAQ52277_aa1	Homo sapiens	VERT- FK506 binding protein (FKBP12A) cDNA.	159	100
1460	AAU20255	Homo sapiens	HUMA- Human novel endocrine antigen, SEQ ID No 312.	104	76
1460	ABB17663	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 6320.	94	77
1460	AAO02331	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 16223.	88	61
1461	AAM65951	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26257.	97	57
1461	AAM53568	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25673.	97	57
1461	AAU83199	Homo sapiens	ZYMO Novel secreted protein Z891639G1P.	96	38
1463	gi555687	Homo sapiens	topoisomerase-related function protein	514	75
1463	gi5139669	Homo sapiens	LAK-1	468	75
1463	gi21430468	Drosophila melanogaster	LP06848p	332	51
1464	AAY91421	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 7 SEQ ID NO:142.	109	35
1464	AAY91396	Homo sapiens	HUMA- Human secreted protein	109	35

161

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			sequence encoded by gene 7 SEQ ID NO:117.		
1464	AAV91352	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 7 SEQ ID NO:73.	109	35
1465	AAU15978	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 931.	575	100
1465	AAU15958	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 911.	575	100
1465	gi16041675	Homo sapiens	joined to JAZF1	575	100
1466	AAO01502	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15394.	173	66
1466	gi10947038 refNP_065209.1	Homo sapiens	ankyrin 1, isoform 1; ankyrin-1, erythrocytic; ankyrin-R	74	28
1466	gi10947036 refNP_065208.1	Homo sapiens	ankyrin 1, isoform 4; ankyrin-1, erythrocytic; ankyrin-R	74	28
1467	gi19354550	Mus musculus	similar to src homology three (SH3) and cysteine rich domain	842	91
1467	AAU17352	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 917.	361	98
1467	gi1799566	Mus musculus	stac	302	44
1468	gi13506771	Mus musculus	structural protein FBF1	767	74
1468	gi7549210	Babesia bigemina	200 kDa antigen p200	213	29
1468	gi1747	Oryctolagus cuniculus	trichohyalin	191	30
1469	gi11345048	Homo sapiens	SCAN domain-containing protein 2	86	32
1469	gi11320940	Homo sapiens	SCAND2	86	32
1469	gi14210722	Tupaia herpesvirus	t41	86	30
1470	AAV88278	Homo sapiens	MILL- Human TANGO 188 protein.	1442	100
1470	gi14336711	Homo sapiens	similar to C. Elegans protein F17C8.5	1442	100
1470	AAA39947_aal	Homo sapiens	MILL- Human TANGO 188 cDNA.	1438	99
1471	AAE10204	Homo sapiens	HYSE- Human bone marrow derived contig protein, SEQ ID NO: 69.	71	44
1471	AAA23458_aal	Homo sapiens	ALPH- cDNA encoding human secreted protein vp15_1, SEQ ID NO:71.	67	46
1471	AAB80228	Homo sapiens	GETH Human PRO269 protein.	67	46
1472	AAB88433	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0210.	136	86
1472	AAB95155	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17188.	136	86
1472	AAE01745	Homo sapiens	HUMA- Human gene 2 encoded secreted protein HOGCS52 variant, SEQ ID NO:160.	136	86
1473	gi9294201	Arabidopsis thaliana	disease resistance protein	70	24
1474	AAE19157	Homo sapiens	THOR/ Human kinase polypeptide (PKIN-15).	631	98
1474	AAM79131	Homo sapiens	HYSE- Human protein SEQ ID NO	494	72



Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			1793.		
1474	AAW19920	Homo sapiens	REGC Human Ksr' (kinase suppressor of Ras).	494	72
1475	AAD12609_aal	Homo sapiens	SAGA Human protein having hydrophobic domain encoding cDNA clone HP03974.	657	73
1475	AAO14199	Homo sapiens	INCY- Human transporter and ion channel TRICH-16.	657	73
1475	AAE06614	Homo sapiens	SAGA Human protein having hydrophobic domain, HP03974.	657	73
1476	gi13905246	Mus musculus	RIKEN cDNA 2410024K20 gene	71	34
1476	gi17505208 refNP_081629.1	Mus musculus	CD2 antigen (cytoplasmic tail) binding protein 2; 1500011B02Rik	71	34
1477	gi806491	Rattus norvegicus	guanylyl cyclase	140	65
1477	gi2648066	Canis familiaris	guanylate cyclase E	118	55
1477	gi2623074	Bos taurus	rod outer segment guanylate cyclase precursor	116	55
1478	gi2065210	Mus musculus	Pro-Pol-dUTPase polyprotein	585	73
1478	gi18676710	Homo sapiens	FLJ00254 protein	408	69
1478	AAO04042	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 17934.	392	75
1479	AAU05396	Homo sapiens	GEHO Human titin (connectin) protein sequence.	208	29
1479	gi1212992	Homo sapiens	Protein sequence and annotation available soon via Swiss-Prot; available at present via e-mail from LABEIT@EMBL-Heidelberg.DE	208	29
1479	gi17066105	Homo sapiens	Titin	208	29
1480	AAV44685_aal	Homo sapiens	TEXA Osteoclast inhibitor protein, OIP-1, coding sequence.	94	41
1480	AAB35287	Homo sapiens	UROG- Human stem cell antigen-2.	94	41
1480	AAV99709	Homo sapiens	REGC Human stem cell antigen-2, hSCA-2.	94	41
1481	AAB57094	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1672.	122	100
1481	gi32672	Homo sapiens	interferon alpha/beta receptor	122	100
1481	AAQ49625_aal	Homo sapiens	EUBI- Human interferon receptor extracellular domain coding sequence.	118	96
1482	AAD17516_aal	Homo sapiens	SENO- Human taste receptor, hT1R1 cDNA coding sequence.	890	94
1482	ABB77319	Homo sapiens	INCY- Human G-protein coupled receptor SEQ ID NO 3.	890	94
1482	AAE10372	Homo sapiens	SENO- Human taste receptor, hT1R1 protein.	890	94
1483	gi18376312	Neurospora crassa	related to SSD1 protein	109	39
1483	gi2645173	Schizosaccharomyces pombe	sts5+	99	42
1483	gi2459997	Candida albicans	protein phosphatase Ssd1 homolog	99	40
1484	gi18569064 refXP_095378.1	Homo sapiens	similar to 40S RIBOSOMAL PROTEIN S3A (V-FOS TRANSFORMATION EFFECTOR	319	96

163

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			PROTEIN)		
1484	gi 20539276 ref XP_095220.2	Homo sapiens	similar to olfactory receptor MOR145-2	259	94
1484	gi 21295882 gb EAA08027.1	Anopheles gambiae str. PEST	agCP1347	68	32
1485	ABB11761	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2131.	197	36
1485	gi930259	Woolly monkey sarcoma virus	reverse transcriptase (476 AA)	148	33
1485	gi18076262	porcine endogenous retrovirus	Pol protein	147	38
1486	AAM74887	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 35193.	172	100
1486	AAM62085	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 34190.	172	100
1486	gi152661	Plasmid pSB24.2	neomycin resistance protein	75	26
1487	gi12653493	Homo sapiens	Similar to brain acid-soluble protein 1	75	34
1487	gi17428832	Ralstonia solanacearum	PROBABLE AVRBS3-LIKE PROTEIN	75	33
1487	gi7329672	Arabidopsis thaliana	phosphatidate cytidyltransferase-like protein	72	46
1488	AAU74754	Homo sapiens	INCY- Human protease PRTS-14 protein sequence.	2042	83
1488	AAU74752	Homo sapiens	INCY- Human protease PRTS-12 protein sequence.	476	39
1488	gi11935122	Mus musculus	papilin	431	40
1489	gi 17543712 ref NP_499976.1	Caenorhabditis elegans	Y55F3C.8.p	72	32
1489	gi 20344600 ref XP_109579.1	Mus musculus	RIKEN cDNA 4933431K05	70	30
1489	gi 11692798 gb AAG40002.1 AF320125.1	Xenopus laevis	ataxia telangiectasia and Rad3-related protein	69	26
1490	AAB95817	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18817.	256	63
1490	ABB06369	Homo sapiens	BODE- Human neurogenesis related protein 12 SEQ ID NO:2.	173	64
1490	AAB44394	Homo sapiens	HUMA- Gene 10 encoded human secreted protein fragment as BLASTX query sequence.	83	66
1491	gi438795	Mus musculus	serotonin 1A receptor	73	26
1491	gi1066326	Mus musculus	serotonin 1A receptor	72	26
1491	gi 438795 gb AAA16850.1	Mus musculus	serotonin 1A receptor	73	26
1492	gi16198083	Drosophila	LD29875p	87	33

164

Table 2

SEQ ID. NO:	Accession No.	Species	Description	Score	% Identity
		melanogaster			
1492	gi2327063	Pneumocystis carinii f. sp. carinii	protease 1	75	34
1492	gi20420	Prunus dulcis	extensin	75	34
1493	AAG67087	Homo sapiens	SHAN- Human ATP-dependent serine protein hydrolase 13.	106	67
1493	AAM76636	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 36942.	103	68
1493	AAM63822	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 35927.	103	68
1494	AAV31225	Homo sapiens	AVET Human RNA helicase p135 protein.	73	38
1494	gi3123906	Homo sapiens	pre-mRNA splicing factor	73	38
1494	gi13278975	Homo sapiens	pre-mRNA splicing factor similar to S. cerevisiae Prp16	73	38
1495	gi17568307 refNP_509837.1	Caenorhabditis elegans	collagen	74	35
1496	gi2065210	Mus musculus	Pro-Pol-dUTPase polypeptide	410	81
1496	gi10834720 gb AAG23790.1 AF258587.1	Homo sapiens	PP565	301	77
1496	gi6753924 refNP_034374.1	Mus musculus	Friend virus susceptibility 1	127	37
1497	gi20901968	Caenorhabditis elegans	C. elegans RPL-36 protein (corresponding sequence F37C12.4)	71	34
1497	gi17554754 refNP_498573.1	Caenorhabditis elegans	Ribosomal protein YL39	71	34
1498	gi5305335	Mycobacterium tuberculosis	proline-rich mucin homolog	102	27
1498	gi330130	human herpesvirus 1	latency associated transcript (LAT) ORF-2	97	37
1498	AAU83682	Homo sapiens	GETH Human PRO protein, Seq ID No 182.	94	30
1499	AAV57937	Homo sapiens	INCY- Human transmembrane protein HTMPN-61.	199	81
1499	AAV36295	Homo sapiens	HUMA- Human secreted protein encoded by gene 72.	151	100
1499	AAG75708	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6472.	141	92
1500	gi21428712	Drosophila melanogaster	SD05267p	165	54
1500	gi20975274	Homo sapiens	skeletrophin	114	40
1500	gi19773434	Mus musculus	skeletrophin	99	52
1501	ABB17830	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 6487.	82	37
1501	AAO12929	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26821.	73	43

165

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1502	gi8778340	Arabidopsis thaliana	F15O4.13	77	39
1503	AAW03515	Homo sapiens	SHKJ Human DOCK180 protein.	144	33
1503	gi1339910	Homo sapiens	DOCK180 protein	144	33
1503	gi13195147	Mus musculus	HCH	129	25
1505	AAM70790	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 31096.	77	53
1505	AAM58316	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 30421.	77	53
1505	gi 21302711 gb EAA14856.1	Anopheles gambiae str. PEST	agCP4916	77	30
1506	AAU75102	Homo sapiens	MYRI- Heat shock protein 8 (Hsp8).	592	79
1506	AAB82535	Homo sapiens	UYCO- Human heat shock protein Hsc70.	592	79
1506	AAE12987	Homo sapiens	SRIV/ Human Hsp70 family homologue, Hsc70.	592	79
1507	ABL53627_aa1	Homo sapiens	GENO- Breast protein-eukaryotic conserved gene 1 (BSTP-ECG1) cDNA.	213	92
1507	ABB75677	Homo sapiens	GENO- Breast protein-eukaryotic conserved gene 1 (BSTP-ECG1) protein.	213	92
1507	AAAY99421	Homo sapiens	GETH Human PRO1433 (UNQ738) amino acid sequence SEQ ID NO:292.	213	92
1508	AAW15565	Homo sapiens	UYJO Human intracellular tyrosine kinase Tnk1-alpha.	79	29
1508	gi233062	Gallus gallus	src downstream region	78	33
1508	gi18376366	Neurospora crassa	related to ribosomal protein S15 precursor (mitochondrial)	72	30
1509	gi 21297482 gb EAA09627.1	Anopheles gambiae str. PEST	agCP15541	68	36
1510	AAM41631	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6562.	127	37
1510	AAM39845	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2990.	127	37
1510	AAM79502	Homo sapiens	HYSE- Human protein SEQ ID NO 3148.	127	37
1511	gi21217669	Mus musculus	myosin IIIA	70	28
1511	gi 21302393 gb EAA14538.1	Anopheles gambiae str. PEST	agCP8799	71	36
1511	gi 20822589 ref XP_140854.1	Mus musculus	similar to myosin IIIA	70	28
1512	gi6911049	Babesia bovis	p9.6.2-like variant erythrocyte surface antigen-1a	82	28
1512	gi6911045	Babesia bovis	p9.6.2 variant erythrocyte surface antigen-1a	82	28
1512	gi6911047	Babesia bovis	p8.4.1 variant erythrocyte surface antigen-1a	81	28

166

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1513	gi10174843	Bacillus halodurans	maltose transport system (permease)	77	25
1513	gi56312	Rattus norvegicus	Gephyrin	76	31
1513	gi4325371	Arabidopsis thaliana	contains similarity to Medicago truncatula N7 protein (GB:Y17613)	74	28
1514	AAAY14196	Homo sapiens	TAKE/ T cell receptor zeta chain protein sequence.	95	100
1514	gi623042	Homo sapiens	T-cell receptor zeta chain	95	100
1514	gi4960202	Sus scrofa	CD3 zeta chain	95	100
1515	ABB07508	Homo sapiens	INCY- Human aminoacyl tRNA synthetase (ATRS) polypeptide (ID: 7474756CD1).	726	100
1515	AAB43670	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1115.	604	82
1515	gi1464742	Homo sapiens	threonyl-tRNA synthetase	604	82
1516	gi21109348	Xanthomonas axonopodis pv. citri str. 306	cytochrome B561	77	29
1516	gi21114046	Xanthomonas campestris pv. campestris str. ATCC 33913	cytochrome B561	76	28
1516	gi21243760  ref NP_643342.1	Xanthomonas axonopodis pv. citri str. 306	cytochrome B561	77	29
1517	ABB11450	Homo sapiens	HYSE- Human neurotoxin homologue, SEQ ID NO:1820.	119	33
1517	gi8809770	Mus musculus	Ly-6I.1	94	30
1517	gi8809768	Mus musculus	lymphocyte antigen LY6I precursor	94	30
1519	gi59977 em b CAA78662.1	Human endogenous retrovirus	tripartite fusion transcript PLA2L	171	67
1519	gi17826947  dbj BAB79287.1	Pseudomonas sp. ND137	beta-1,4-xylanase	73	34
1519	gi21232680  ref NP_638597.1	Xanthomonas campestris pv. campestris str. ATCC 33913	ribonuclease PH	72	30
1520	AAM78023	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 38329.	190	100
1520	AAM65326	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 37431.	190	100
1520	gi13447468	Emericella nidulans	FH1/FH2 protein homolog	121	49
1522	AAG81417	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:352.	287	100
1523	AAAY90349	Homo sapiens	SMIK Human fatty acid synthase (FAS) protein sequence.	158	85
1523	AAB43871	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1316.	158	85

167

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1523	gi915392	Homo sapiens	fatty acid synthase	158	85
1525	AAG03819	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7900.	93	100
1525	gi1311466	Homo sapiens	24-kDa subunit of Complex I	93	100
1525	gi188852	Homo sapiens	NADH-ubiquinone reductase	93	100
1526	AAD02855_aa1	Homo sapiens	SUKA Human platelet membrane glycoprotein VI (GPVI) cDNA.	73	31
1526	AAB49403	Homo sapiens	MERE Human glycoprotein VI mature protein.	73	31
1526	AAB61257	Homo sapiens	MILL- Mature human TANGO 268 protein.	73	31
1527	gi17864896	Mus musculus	protocadherin 18 precursor	81	31
1527	gi15980222	Yersinia pestis	aconitate hydratase 1	79	30
1527	gi12248353	Fasciola hepatica	NADH dehydrogenase subunit 5	75	56
1528	gi2440214	Trypanosoma brucei brucei	invariant surface glycoprotein 100	83	28
1528	gi10567463	Rhizobium rhizogenes	probable virB1 gene	78	22
1529	gi2231279	Porcine reproductive and respiratory syndrome virus	envelope protein	66	31
1530	gi 199851 gb AAA39757.1	Mus musculus	pol protein	257	42
1530	gi 1498648 gb AAB06450.1	Mus musculus	Gag-Pol polyprotein	257	42
1530	gi 331995 gb AAB03091.1	AKV murine leukemia virus	gag-pol polyprotein (tag amber codon at 2250-2252 inserts Gln in Mo-MuLV)	257	42
1533	gi435698	Homo sapiens	CD44SP	136	100
1533	AAV63461_aa1	Homo sapiens	GEHO Human CD44 antigen cDNA.	130	100
1533	AAT14724_aa1	Homo sapiens	GEHO Human haematopoietic CD44 cDNA clone CD44.5.	130	100
1534	gi2622165	Methanothermobacter thermotrophicus str. Delta H	acetyltransferase	71	29
1534	gi 15679078 ref NP_276195.1	Methanothermobacter thermotrophicus	acetyltransferase	71	29
1535	gi7777	Drosophila melanogaster	protein H	73	28
1535	gi457146	Plasmodium yoelii	rhopty protein	73	38
1535	gi13195258	Plasmodium yoelii yoelii	235 kDa rhopty protein	73	38
1536	ABB09740	Homo sapiens	BODE- Amino acid sequence of human protein phosphatase 11.66.	132	43
1536	gi 20830386 ref XP_1456	Mus musculus	similar to importin alpha 1b	72	35

168

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	42.1				
1537	gi14039907	Rattus norvegicus	cytochrome P450 monooxygenase CYP2T1	353	39
1537	gi2920650	Mus musculus	cytochrome P450 CYP2B19	275	44
1537	gi2353336	Capra hircus	cytochrome P450	271	31
1538	AAU83175	Homo sapiens	ZYMO Novel secreted protein Z874015G4P.	282	100
1538	gi6714803	Streptomyces coelicolor A3(2)	integral membrane protein.	77	26
1539	gi12963397	Prunus x yedoensis	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	74	32
1539	gi466436	Saccharomyces cerevisiae	BOI1	69	31
1539	gi5833897	Besleria affinis	ribulose 1,5-bisphosphate carboxylase large subunit	69	31
1542	AAV32193	Homo sapiens	INCY- Human receptor molecule (REC) encoded by Incyte clone 044150.	73	26
1542	gi7576677	Helicobacter pylori	IceA1	72	44
1542	gi20841498 ref XP_131541.1	Mus musculus	similar to MUF1 protein	73	26
1546	gi14581448	Homo sapiens	FSHD Region Gene 2 protein	73	42
1546	gi15982852	Arabidopsis thaliana	AT5g66850/MUD21_11	71	34
1546	gi14581448 gb AAK21977.1	Homo sapiens	FSHD Region Gene 2 protein	73	42
1547	gi18676660	Homo sapiens	FLJ00229 protein	192	92
1547	AAU21409	Homo sapiens	HUMA- Human novel foetal antigen, SEQ ID NO 1653.	179	100
1547	AAM42128	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 7059.	114	53
1548	AAG64494	Homo sapiens	SHAN- Human natriuretic peptide receptor 18.	539	100
1548	gi18676710	Homo sapiens	FLJ00254 protein	268	77
1548	AAB28764	Homo sapiens	HUMA- Sequence homologous to protein fragment encoded by gene 21.	249	72
1549	AAB67055	Homo sapiens	INCY- Human immune response molecule (IMUN) protein SEQ ID NO: 9.	606	82
1549	AAO01862	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15754.	404	72
1549	gi6753924 ref NP_034374.1	Mus musculus	Friend virus susceptibility 1	213	36
1550	gi190129	Homo sapiens	70kDa peroxisomal membrane protein	92	100
1550	gi825711	Homo sapiens	70kD peroxisomal integral membrane protein	92	100
1550	gi220862	Rattus norvegicus	PMP70	89	94
1551	AAM69543	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ	228	100

169  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			ID NO: 29849.		
1551	AAM57148	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 29253.	228	100
1551	AAB93944	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13960.	94	57
1552	gi4884924	Rangiferine herpesvirus 1	glycoprotein C	75	34
1552	gi 18556240 ref XP_067628.2	Homo sapiens	similar to Salivary glue protein SGS-3 precursor	78	30
1552	gi 4884924 gb AAD31876.1	Rangiferine herpesvirus 1	glycoprotein C	75	34
1553	gi 2193870 dbj BAA20419.1	Mus musculus	reverse transcriptase	176	35
1553	gi 2731767 gb AAC53542.1	Mus musculus	endonuclease/reverse transcriptase	176	35
1554	ABB08776	Homo sapiens	BODE- Human neuregulin 55 SEQ ID NO 2.	75	29
1554	AAM92816	Homo sapiens	HUMA- Human digestive system antigen SEQ ID NO: 2165.	71	29
1554	gi 6322838 ref NP_012911.1	Saccharomyces cerevisiae	Protein required for cell viability; Ykl014cp	70	27
1555	gi7528184	Drosophila melanogaster	bicoid-interacting protein BIN3	78	28
1555	gi15292595	Drosophila melanogaster	SD09926p	78	28
1555	gi4514620	Mus musculus	Ror2	71	24
1557	ABA91504_aal	Homo sapiens	EYEE- Human epidermal growth factor receptor precursor cDNA.	144	93
1557	AAF85332_aal	Homo sapiens	NOVS Nucleotide sequence of wild type EGFR1.	144	93
1557	AAM50768	Homo sapiens	EYEE- Human epidermal growth factor receptor precursor.	144	93
1558	AAB99950	Homo sapiens	SHAN- Human alkylated-DNA-protein cysteine methyltransferase 14.	221	100
1558	AAU16267	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1220.	221	100
1558	ABB11507	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:1877.	183	97
1559	gi14599730	Spachea corraeae	maturase	71	28
1559	gi14599648	Blepharandra heteropetala	maturase	71	30
1559	gi14599673	Galphimia gracilis	maturase	70	28
1560	gi2323287	multiple sclerosis associated retrovirus	polyprotein	340	83
1560	gi 13310191	multiple	recombinant envelope protein	260	70



170

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	gb AAK18189.1 AF331500_1	sclerosis associated retrovirus element			
1560	gi 21103962 gb AAM33141.1	Homo sapiens	enverin-2	248	84
1561	AAB94698	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15680.	107	95
1561	AAU18480	Homo sapiens	HUMA- Human endocrine polypeptide SEQ ID No 435.	107	95
1561	ABB10288	Homo sapiens	HUMA- Human cDNA SEQ ID NO: 596.	107	95
1562	gi969078	Drosophila melanogaster	S-adenosylhomocysteine hydrolase	73	26
1562	gi21064553	Drosophila melanogaster	RE58316p	73	26
1562	AAM41205	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6136.	72	30
1563	gi1778844	Dictyostelium discoideum	LimA	71	34
1563	gi 20985456 ref XP_142111.1	Mus musculus	similar to actin beta chain - human	75	36
1563	gi 1778844 gb AAB40929.1	Dictyostelium discoideum	LimA	71	34
1564	gi 9507757 ref NP_061423.1	Plasmid F	resolvase	507	91
1564	gi 148589 gb AAA24900.1	Plasmid F	Protein D	507	91
1564	gi 10955295 ref NP_052636.1	Escherichia coli	resolvase	501	90
1565	gi7649370	Arabidopsis thaliana	guanine nucleotide-exchange-like protein	77	38
1565	gi1674160	Mycoplasma pneumoniae	involved in cytodherence, see: MPN142	71	35
1565	gi 15229258 ref NP_189916.1	Arabidopsis thaliana	guanine nucleotide-exchange - like protein	77	38
1566	gi1799600	SwissProt Accession Number P31458	similar to	1051	99
1566	gi13814506	Sulfolobus solfataricus	Mandelate racemase /muconate lactonizing enzyme related protein (MR/MLE)	286	35
1566	gi10640034	Thermoplasma acidophilum	starvation-sensing protein rspA related protein	270	35
1567	gi13359972	Escherichia coli O157:H7	acridine efflux pump	573	98
1567	gi1773144	Escherichia coli	probable transmembrane protein AcrE	573	98

171

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1567	gi532311	Escherichia coli	114 kDa protein	573	98
1569	gi8918871	YccA of plasmid Collb-P9] [Plasmid F	96 pct identical to gp:AB021078_30	288	98
1569	gi17136976 refNP_477026.1	Drosophila melanogaster	repo-P1; Antibody RK2	71	33
1569	gi6502544 gb AAFI4351.1 AF110198.1	Glomus intraradices	homeobox protein HB1	70	31
1570	gi13363792	Escherichia coli O157:H7	zinc-transporting ATPase	410	87
1570	gi466605	Escherichia coli	No definition line found	410	87
1570	gi12518128	Escherichia coli O157:H7 EDL933	zinc-transporting ATPase	410	87
1571	AAU83186	Homo sapiens	ZYMO Novel secreted protein Z887014G7P.	1006	100
1571	gi7248459	Zea mays	arabinogalactan protein	85	29
1571	gi3513742	Arabidopsis thaliana	contains similarity to Zea mays embryogenesis transmembrane protein (GB:X97570)	82	35
1572	gi12597465	Caenorhabditis elegans	CED-1	72	44
1572	gi19571666	Caenorhabditis elegans	similar to EGF-like domain	72	44
1572	gi4883938	Drosophila melanogaster	laminin alpha1,2	67	31
1573	ABB12490	Homo sapiens	HYSE- Human bone marrow expressed protein SEQ ID NO: 329.	106	38
1574	gi1478205	Mus musculus	PNG protein	75	41
1574	AAM40148	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3293.	69	56
1574	AAM79341	Homo sapiens	HYSE- Human protein SEQ ID NO 2987.	69	35
1576	gi20882651 refXP_123303.1	Mus musculus	ATPase, class 2, member b	234	91
1576	gi7656918 refNP_056620.1	Mus musculus	ATPase, class 2, member b; ATPase 9B, class II; ATPase 9B, p type	234	91
1577	gi18143418	Alteromonas sp. O-7	chitinase A	77	39
1577	gi15426105	Leishmania major	probable surface antigen protein	75	24
1578	gi19702241	Homo sapiens	rabconnectin	439	93
1578	gi7452946	Homo sapiens	X-like 1 protein	132	41
1578	gi1279384	Drosophila melanogaster	X	109	29
1580	AAE20337	Homo sapiens	HUMA- Human B7-H11 protein mature extracellular domain.	122	23
1580	AAE20336	Homo sapiens	HUMA- Human B7-H11 protein extracellular domain.	122	23

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1580	gi2062702	Homo sapiens	butyrophilin	122	23
1581	AAE18640	Homo sapiens	INCY- Human G-protein coupled receptor (GCRC-1).	70	35
1581	gi18369751	Oryza sativa	ethylene responsive protein	70	50
1581	gi15217292	Oryza sativa] [Oryza sativa (japonica cultivar-group)	Putative AP2 domain containing protein	70	50
1583	gi6468047	Homo sapiens	Kruppel-like factor	85	73
1583	gi5916096	Homo sapiens	Kruppel-like factor LKLF	85	73
1583	gi4583418	Homo sapiens	Kruppel-like zinc finger transcription factor	85	73
1585	gi2570021	Homo sapiens	paired box containing transcription factor	77	37
1585	gi3115988	Homo sapiens	dJ394P21.1 (PAX-7)	77	37
1585	gi2570015	Homo sapiens	alternative	77	37
1586	gi7861533	Rattus norvegicus	retina specific protein PAL	72	43
1586	gi20977028	Xenopus laevis	mitotic phosphoprotein 39	72	34
1586	AAB58458	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 796.	68	39
1587	gi5901864	Drosophila melanogaster	BcDNA.LD27873	81	24
1587	gi15458514	Streptococcus pneumoniae R6	Pneumococcal histidine triad protein D precursor	78	27
1587	gi5042400	Homo sapiens	NFI-X3=transcription factor [AA	75	30
1592	gi4210501	Homo sapiens	BC85722 1	253	61
1592	gi14794910	Homo sapiens	capicua protein	253	61
1592	gi14794914	Mus musculus	capicua protein	253	61
1593	gi 8131854 gb AAF73108.1 AF147956.1	Trypanosoma cruzi	antigen JL8	69	34
1595	gi18892729	Pyrococcus furiosus DSM 3638	3-hydroxyisobutyrate dehydrogenase	70	27
1595	gi 20847046 ref XP_136621.1	Mus musculus	similar to Transcription factor BTF3 (RNA polymerase B transcription factor 3)	70	28
1595	gi 18977088 ref NP_578445.1	Pyrococcus furiosus DSM 3638	3-hydroxyisobutyrate dehydrogenase	70	27
1597	AAU83621	Homo sapiens	GETH Human PRO protein, Seq ID No 60.	151	42
1597	AAO05826	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 19718.	146	83
1597	AAM41346	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6277.	102	46
1598	AAM79503	Homo sapiens	HYSE- Human protein SEQ ID NO 3149.	80	35
1598	AAM78519	Homo sapiens	HYSE- Human protein SEQ ID NO 1181.	80	35
1598	gi18676526	Homo sapiens	FLJ00160 protein	80	35
1599	gi2149640	Arabidopsis	Argonaute protein	72	33

173

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		<i>thaliana</i>			
1599	gi15027491	respiratory syncytial virus	glycoprotein	71	32
1599	gi15221177 refNP_175274.1	<i>Arabidopsis thaliana</i>	leaf development protein Argonaute	72	33
1601	gi17130010	<i>Nostoc</i> sp. PCC 7120	WD-40 repeat protein	136	28
1601	gi1653631	<i>Synechocystis</i> sp. PCC 6803	beta transducin-like protein	131	26
1601	gi17135261	<i>Nostoc</i> sp. PCC 7120	WD-40 repeat protein	115	27
1602	gi1103853	<i>Rattus norvegicus</i>	rHAP1-A	89	33
1602	gi1103851	<i>Rattus norvegicus</i>	huntingtin associated protein	89	33
1602	gi14579673	<i>Takifugu rubripes</i>	pericentriolar material 1 protein	87	30
1603	gi537446	<i>Arabidopsis thaliana</i>	AtHSP101	75	31
1603	gi12324908	<i>Arabidopsis thaliana</i>	heat shock protein 101; 13093-16240	75	31
1603	gi6715468	<i>Arabidopsis thaliana</i>	heat shock protein 101	75	31
1604	gi2190531	<i>Vibrio cholerae</i>	methyl accepting chemotaxis protein	71	26
1604	gi9657614	<i>Vibrio cholerae</i>	hemolysin secretion protein HylB	71	26
1604	gi9655306	<i>Vibrio cholerae</i>	heat shock protein GrpE	70	35
1605	gi3912936	<i>Geobacillus stearothermophilus</i>	ornithine carbamoyltransferase	68	31
1606	gi8797	<i>Drosophila melanogaster</i>	CYS3HIS finger protein	678	51
1606	gi15291975	<i>Drosophila melanogaster</i>	LD33756p	617	65
1606	gi6967181	<i>Homo sapiens</i>	c399E4.1 (similar to D.melanogaster unkempt protein.)	549	75
1607	gi21301783 gb EAA13928.1	<i>Anopheles gambiae</i> str. PEST	agCP8730	72	35
1607	gi21361276 refNP_006075.2	<i>Homo sapiens</i>	interferon-stimulated transcription factor 3, gamma (48kD); interferon-stimulated gene factor 3, gamma subunit (48 kD)	68	29
1609	gi2661094	<i>Spinacia oleracea</i>	cold acclimation protein	76	32
1612	gi1780975 emb CAA71418.1	Human endogenous retrovirus K	gag protein	312	34
1612	gi5802810 gb AAD51791.1	<i>Homo sapiens</i>	Gag-Pro-Pol protein	309	34
1612	gi887448 emb CAA51306.1	Human endogenous retrovirus	gag	309	34

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1613	AAO13889	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 27781.	73	42
1614	gi11065727	Homo sapiens	dJ493F7.1 (similar to murine BET3)	347	100
1614	gi2791806	Mus musculus	bet3	253	69
1614	gi13277654	Mus musculus	Bet3 homolog (S. cerevisiae)	253	69
1615	gi1122901	Saccharomyces cerevisiae	MSP8	77	20
1615	gi825546	Saccharomyces cerevisiae	Cat8p	77	20
1615	gi17978563	Xenopus laevis	Sp1-like zinc-finger protein XSPR-1	75	40
1616	AAY02536	Homo sapiens	ICOS- Human ICAM-6 protein sequence.	458	98
1616	gi12248907	Homo sapiens	TCAM-1	458	98
1616	gi4579740	Rattus norvegicus	testicular cell adhesion molecule 1 (TCAM1)	366	76
1617	AAM67067	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27373.	271	64
1617	AAM54664	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26769.	271	64
1617	AAM56747	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 28852.	229	69
1618	gi5802814	Homo sapiens	Gag-Pro-Pol-Env protein	532	52
1618	gi1780973	Human endogenous retrovirus K	pol protein	531	52
1618	gi5802821	Homo sapiens	Gag-Pro-Pol protein	531	52
1619	gi2769587	Mus musculus	STOP protein	662	86
1619	gi1370291	Rattus norvegicus	STOP protein	662	92
1619	gi3287265	Rattus norvegicus	E-STOP protein	662	92
1620	AAM65980	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26286.	266	100
1620	AAM53601	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25706.	266	100
1620	gi 20270271 ref NP_62082.1	Mus musculus	RIKEN cDNA 1190017O12	198	80
1621	gi11862941	Mus musculus	DDM36E	74	33
1621	gi11862939	Mus musculus	DDM36	74	33
1621	gi7650186	Mus musculus	neighbor of Punc e11 protein	73	33
1622	gi3157464	Thermus sp. A4	integral membrane protein	74	38
1623	gi 59977 emb CAA78662.1	Human endogenous retrovirus	tripartite fusion transcript PLA2L	129	82
1623	gi 20161147 dbj BAB90075.1	Oryza sativa (japonica cultivar-group)	VsaA -like protein	88	32
1623	gi 17864474	Drosophila	domino	87	41

175  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	ref NP_524833.1	melanogaster			
1626	AAO00498	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 14390.	99	43
1627	gi14041733	Xenorhabdus nematophila	XptA2 protein	70	23
1627	gi15641593 ref NP_231225.1	Vibrio cholerae	catalase	69	23
1628	gi19888204	Methanopyrus kandleri AV19	Site-specific DNA methylase	80	27
1628	gi6358691	Simian immunodeficiency virus	Pol protein	78	32
1628	gi20094956 ref NP_614803.1	Methanopyrus kandleri AV19	Site-specific DNA methylase	80	27
1629	AAB07704	Homo sapiens	INMR Protein encoded by the endogenetic fragment of HERV-W.	594	67
1629	gi8272464	Homo sapiens	gag	594	67
1629	AAB07703	Homo sapiens	INMR Protein encoded by the endogenetic fragment of HERV-W.	590	66
1630	gi32498	Homo sapiens	precursor (AA -23 to 476)	145	100
1630	gi339595	Homo sapiens	triglyceride lipase precursor	145	100
1630	gi386859	Homo sapiens	hepatic lipase	145	100
1631	gi8777465	Rattus norvegicus	cytoplasmic dynein heavy chain	703	77
1631	gi17019507	Tripeustes gratilla	dynein heavy chain isotype 1B	505	53
1631	AAB93815	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13606.	457	71
1632	AAM68837	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29143.	122	48
1632	AAM56460	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 28565.	122	48
1632	gi17861826	Drosophila melanogaster	GM01964p	90	51
1633	gi21300783 gb EAA12928.1	Anopheles gambiae str. PEST	ebiP1105	77	33
1633	gi19880523 gb AAM00372.1 AF368053.1	Bactrocera dorsalis	vitellogenin 1 precursor	68	27
1633	gi21070999 ref NP_065911.1	Homo sapiens	stromal interaction molecule 2 precursor	68	39
1637	gi2323287	multiple sclerosis associated retrovirus	polyprotein	289	91
1637	gi21103962	Homo sapiens	enverin-2	261	82

176  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	gb AAM33141.1				
1637	gi 13310191 gb AAK18189.1 AF331500_1	multiple sclerosis associated retrovirus element	recombinant envelope protein	259	82
1638	AAR58809	Homo sapiens	UYNH Human RPTP-gamma.	86	26
1638	gi292411	Homo sapiens	receptor-type protein tyrosine phosphatase gamma	86	26
1638	gi1263069	Homo sapiens	receptor tyrosine phosphatase gamma	86	26
1639	gi9857054	Leishmania major	possible CG7055 protein	74	27
1639	gi 20853034 ref XP_125962.1	Mus musculus	expressed sequence AI447519	73	35
1639	gi 7008003 dbj BAA90874.1	Mus musculus	transcription factor MAZR	73	35
1640	AAG03810	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7891.	220	95
1640	gi186800	Homo sapiens	ribosomal protein L12	220	95
1640	gi57680	Rattus rattus	ribosomal protein L12	220	95
1641	AAB44286	Homo sapiens	GETH Human PRO1072 (UNQ529) protein sequence SEQ ID NO:303.	1709	100
1641	AAAY41730	Homo sapiens	GETH Human PRO1072 protein sequence.	1709	100
1641	gi14602625	Homo sapiens	PAN2 protein	1709	100
1642	gi20147241	Arabidopsis thaliana	AT5g09850/MYH9_6	74	32
1642	gi14329782	Homo sapiens	dJ1121G12.3 (Novel gene)	72	28
1642	gi 16648730 gb AAL25557.1	Arabidopsis thaliana	AT5g09850/MYH9_6	74	32
1643	gi2952340	Rattus norvegicus	insulin receptor substrate 2	89	31
1643	gi2653351	Bovine herpesvirus type 1.1	product of latency-related gene	83	30
1643	gi4511969	Homo sapiens	insulin receptor substrate-2	82	26
1644	gi9964099	Chlamydia trachomatis	inclusion membrane protein	73	35
1644	gi19171028	Encephalitozoon cuniculi	ATP DEPENDENT DNA BINDING HELICASE (RAD3/XPD SUBFAMILY OF HELICASES)	67	29
1644	gi 9964095 gb AAG09821.1 AF279362_1	Chlamydia trachomatis	inclusion membrane protein	73	35
1646	gi 10863995 ref NP_067011.1	Homo sapiens	clones 23667 and 23775 zinc finger protein	67	42
1647	gi1196425	Homo sapiens	envelope protein	93	39
1647	gi200296	Mus musculus	perlecan	85	26

177

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1647	gi8131894	Homo sapiens	mitofilin	84	27
1648	gi1573040	Haemophilus influenzae Rd	aspartokinase I / homoserine dehydrogenase I (thrA)	73	36
1648	gi8778726	Arabidopsis thaliana	T25N20.14	73	31
1648	gi 16272063 ref NP_438262.1	Haemophilus influenzae Rd	aspartokinase I / homoserine dehydrogenase I (thrA)	73	36
1649	gi295642	Saccharomyces cerevisiae	phospholipase C	79	36
1649	gi7548846	Saccharomyces cerevisiae	delta class phosphoinositide-specific phospholipase C homolog	77	36
1649	gi161104	Schistosoma mansoni	engrailed-like homeodomain protein	74	35
1651	gi 13129464 gb AAK13122.1 AC080019_14	Oryza sativa [Oryza sativa (japonica cultivar-group)]	Polyprotein	66	40
1652	AAG81446	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:410.	249	100
1652	gi18032212	Homo sapiens	histone acetyltransferase MOZ2	89	34
1652	AAR34936	Homo sapiens	UYJO CENP-B.	77	35
1653	gi20145484	Bos taurus	SCO-spondin	71	29
1655	AAM86382	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:13975.	129	55
1655	ABB03887	Homo sapiens	HUMA- Human musculoskeletal system related polypeptide SEQ ID NO 1834.	118	62
1655	AAM75964	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 36270.	85	56
1659	gi38035	Homo sapiens	p25 protein	110	45
1659	gi330915	Equine herpesvirus 1	IR4 protein	99	28
1659	gi156606	Chironomus tentans	SpId	84	30
1660	gi9654641	Vibrio cholerae	3-deoxy-D-manno-octulosonic-acid transferase	84	23
1660	gi 20835446 ref XP_144409.1	Mus musculus	similar to STARP antigen	73	25
1660	gi 15596880 ref NP_250374.1	Pseudomonas aeruginosa	probable sugar aldolase	72	26
1661	gi4062318	Escherichia coli	Heat-responsive regulatory protein	79	36
1661	gi976025	Escherichia coli	HrsA	79	36
1661	gi1786951	Escherichia coli K12	protein modification enzyme, induction of ompC	79	36
1662	AAM68588	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 28894.	155	100
1662	AAM56212	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID	155	100



Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO: 28317.		
1662	gi3845169	Plasmodium falciparum 3D7	phosphatase (acid phosphatase family)	66	52
1663	AAG89215	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 335.	218	100
1663	gi20070921	Mus musculus	RIKEN cDNA 2410008M22 gene	130	55
1663	AAR77602	Homo sapiens	FORS/ Human circulating cytokine CC-1 C-terminal fragment.	92	44
1664	AAE18212	Homo sapiens	CURA- Human MOL4 protein.	75	47
1664	AAM00966	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 442.	72	35
1665	AAB92828	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11365.	74	93
1665	AAG63852	Homo sapiens	INCY- Amino acid sequence of human GTPase activating protein GTPAP2.	74	93
1665	AAG63851	Homo sapiens	INCY- Amino acid sequence of human GTPase activating protein GTPAP1.	74	93
1666	AAM72897	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 33203.	135	65
1666	AAM60268	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 32373.	135	65
1666	gi4007097	Homo sapiens	dJ1118D24.2 (60S Ribosomal Protein L10 LIKE)	135	65
1667	gi212267	Gallus gallus	cartilage link protein	917	49
1667	gi2010	Sus scrofa	link protein precursor (AA -15 to 339)	913	51
1667	gi459439	Equus caballus	link protein	910	51
1668	gi10443237	Mus musculus	splicing factor 3a, subunit 2	276	36
1668	gi396743	Podocoryne carnea	Pod-EPPT	276	30
1668	gi294131	Plasmodium falciparum	circumsporozoite protein	266	22
1669	AAM49641	Homo sapiens	BOEH Human tumour-associated antigen B345 protein SEQ ID NO 4.	132	65
1669	AAU12252	Homo sapiens	GETH Human PRO5773 polypeptide sequence.	132	65
1669	AA91592	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 6 SEQ ID NO:265.	132	65
1670	gi4835383	Homo sapiens	alias DLC1	226	47
1670	gi4704343	Homo sapiens	alias DLC1; candidate tumor suppressor gene	226	47
1670	gi155627	Acanthamoeba castellanii	myosin I heavy chain	118	42
1671	ABB12490	Homo sapiens	HYSE- Human bone marrow expressed protein SEQ ID NO: 329.	237	88
1671	gi6002932	Streptomyces fradiae	glycosyl transferase	67	35
1671	gi9634613 refNP_038150.1	Human papillomavirus type 69	L1	65	39
1672	gi13938013	Homo sapiens	Similar to RIKEN cDNA 2610509G12 gene	333	66

179

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1672	gi2388970	Schizosaccharomyces pombe	tat-binding homolog 7, AAA ATPase family protein	235	41
1672	gi6850321	Arabidopsis thaliana	Contains similarity to YTA7 ATPase gene from Saccharomyces cerevisiae gb X81072, and contains Bromodomain PF 00439, AAA PF 00004, and Sigma-54 PF 00158 transcription factor domains.	214	40
1673	gi11066113	Drosophila melanogaster	Misexpression suppressor of ras 4	71	29
1673	gi 20829387 ref XP_129540.1	Mus musculus	RIKEN cDNA 4930455F23	77	27
1673	gi 17647635 ref NP_523775.1	Drosophila melanogaster	Misexpression suppressor of ras 4	71	29
1674	gi 20535935 ref XP_115787.1	Homo sapiens	similar to splicing coactivator subunit SRm300; RNA binding protein; AT-rich element binding factor	75	37
1674	gi 17544226 ref NP_500151.1	Caenorhabditis elegans	Y76B12C.4.p	72	34
1674	gi 17559826 ref NP_505799.1	Caenorhabditis elegans	sepB domain	70	26
1675	gi5708067	Oryctolagus cuniculus	hyperpolarization activated cation channel	99	27
1675	gi402558	Canis familiaris	mucin	98	27
1675	gi10636484	Homo sapiens	polyglutamine-containing protein	96	26
1676	AAM95365	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4023.	73	26
1676	AAB56709	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1287.	72	34
1676	gi1881288	Bacillus subtilis	FUNCTION UNKNOWN, SIMILAR PRODUCT IN E.COLI, H. INFLUENZAE AND NEISSERIA MENINGITIDIS.	71	30
1677	gi 15892512 ref NP_360226.1	EC:2.7.7.41 [Rickettsia conorii	phosphatidate cytidyltransferase	65	34
1679	gi14231	Saccharomyces cerevisiae	NADH dehydrogenase (ubiquinone)	75	31
1679	gi805022	Saccharomyces cerevisiae	Ndi1p	73	31
1679	gi1353352	Chlamydomonas reinhardtii	alanine aminotransferase	70	27
1680	gi1805421	Bacillus subtilis	surfactin production	77	36
1680	gi396482	Bacillus subtilis	srfA2	77	36
1680	gi516360	Bacillus subtilis	surfactin synthetase	77	36
1681	AAG64494	Homo sapiens	SHAN- Human natriuretic peptide receptor 18.	156	80
1681	AAE16275	Homo sapiens	INCY- Human kinase PKIN-21 protein.	154	73
1681	AAM40599	Homo sapiens	HYSE- Human polypeptide SEQ ID	154	73

180

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO 5530.		
1682	gi2323287	multiple sclerosis associated retrovirus	polyprotein	1646	75
1682	gi2351212 d bj BAA2206 4.1	Friend murine leukemia virus	gag-pol polyprotein (precursor protein)	807	40
1682	gi9626961 ref NP_0579 33.1	Murine leukemia virus	Pr180	802	40
1683	AAM39205	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2350.	457	53
1683	gi3033415	Gibbon ape leukemia virus	gag polyprotein	353	38
1683	gi6524623 gb AAF15097 .1	Phascolarctos cinereus	gag protein	343	38
1684	gi19110438	Homo sapiens	polycystin-1L1	712	98
1684	gi6361629	Periplaneta americana	vitellogenin	81	25
1684	gi3115393	Rana pipiens	guanylate cyclase inhibitory protein	80	35
1686	AAY91542	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 92 SEQ ID NO:215.	212	84
1686	gi1279841	Bos taurus	glycine transporter	72	36
1686	gi19879917	Oryza sativa	acid phosphatase	70	35
1687	gi12056568	Homo sapiens	MSTP063	212	88
1687	gi13539684	Homo sapiens	zinc finger protein 291	212	88
1687	gi12056568 gb AAG479 45.1 AF119 814_1	Homo sapiens	MSTP063	212	88
1689	gi5689766	Homo sapiens	zinc finger 2.2	222	91
1689	AAU16267	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1220.	178	58
1689	AAB99950	Homo sapiens	SHAN- Human alkylated-DNA-protein cysteine methyltransferase 14.	177	60
1690	gi3328880	Chlamydia trachomatis	Protein Export	73	29
1690	gi2832232	Brucella melitensis biovar Abortus	flagellin; FlIC	67	29
1690	gi17984285	Brucella melitensis	FLAGELLIN	67	29
1692	gi4927443	Haemophilus influenzae	hemoglobin/hemoglobin-haptoglobin binding protein	93	80
1692	gi4204775	Haemophilus influenzae	hemoglobin and hemoglobin-haptoglobin binding protein	93	80
1692	gi3647226	Haemophilus influenzae	hemoglobin binding protein	93	80
1694	AAW95631	Homo sapiens	GEMY Homo sapiens secreted protein gene clone hj968 2.	102	100
1694	gi13162186	Homo sapiens	calsyntenin-3 protein	102	100

181  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1695	AAO04205	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 18097.	81	37
1695	gi160180	Plasmodium cynomolgi	circumsporozoite antigen	81	29
1695	gi495522	Plasmodium simiovale	circumsporozoite protein	80	30
1696	AAM80223	Homo sapiens	HYSE- Human protein SEQ ID NO 3869.	252	66
1696	AAM79239	Homo sapiens	HYSE- Human protein SEQ ID NO 1901.	252	66
1696	gi3688394	Homo sapiens	triple LIM domain protein	252	66
1697	gi19887715	Methanopyrus kandleri AV19	Predicted membrane protein	74	28
1698	AAM93184	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2552.	269	87
1698	gi18044066	Mus musculus	RIKEN cDNA 5033406L14 gene	226	76
1698	AAB95302	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17538.	194	78
1699	ABB17279	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 5936.	110	56
1699	AAO13013	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26905.	101	71
1699	gi7650258 gb AAF65960.1 AF207770.1	Hepatitis C virus	polyprotein	74	28
1700	gi12697585	Arabidopsis thaliana	4-(cytidine 5'-phospho)-2-C-methyl-D-erithritol kinase	69	40
1701	gi16740569	Homo sapiens	Similar to thymus expressed gene 3	84	27
1701	gi17940760	Mus musculus	cask-interacting protein 2	79	26
1701	gi17940758	Homo sapiens	cask-interacting protein 1	77	26
1702	gi17385401	Homo sapiens	TPIP alpha lipid phosphatase	234	62
1702	AAU75783	Homo sapiens	INCY- Human protein phosphatase 1 (PP1) protein sequence.	208	57
1702	AAG67638	Homo sapiens	HELI- Amino acid sequence of a human protein.	202	56
1703	AAO07887	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 21779.	246	85
1703	AAO08651	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 22543.	239	83
1703	AAO08732	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 22624.	221	80
1704	AAB94588	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15392.	82	52
1704	gi3288914	Mus musculus	aortic carboxypeptidase-like protein ACLP	82	24
1704	AAM93437	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3074.	81	32
1706	AAM86104	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:13697.	179	100
1706	gi10039425	Equus caballus	ALR protein	120	40
1706	gi20502826	Eimeria maxima	cGMP-dependent protein kinase	115	35
1707	AAM70251	Homo sapiens	MOLE- Human bone marrow	115	78

182

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			expressed probe encoded protein SEQ ID NO: 30557.		
1707	AAM57834	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 29939.	115	78
1707	gi15450860	Arabidopsis thaliana	serine/threonine-protein kinase Mak (male germ cell-associated kinase)-like protein	71	56
1708	gi1620403	Homo sapiens	SF1-Bo isoform	82	41
1708	gi19072991	Hypocrea virens	class III chitinase precursor	82	40
1708	gi18765873	Hypocrea virens	class III chitinase	82	40
1709	AAM52240	Homo sapiens	INCY- Human MFAP4 SEQ ID NO 3.	1384	100
1709	gi790817	Homo sapiens	microfibril-associated glycoprotein 4	1384	100
1709	AAM52239	Homo sapiens	INCY- Human MAG4V SEQ ID NO 1.	1374	100
1710	gi16769882	Drosophila melanogaster	SD07884p	67	27
1710	gi 17545505 ref NP_518907.1	Ralstonia solanacearum	CONSERVED HYPOTHETICAL PROTEIN	66	41
1711	AAU82954	Homo sapiens	ANAD- Human homologue of MPT1 protein target for antifungal compound.	111	27
1711	gi2058326	Homo sapiens	subunit of RNA polymerase II transcription factor TFIID	111	27
1711	gi13559031	Homo sapiens	bA11M20.1 (TATA box binding protein (TBP)-associated factor, RNA polymerase II, C1, 130kD)	108	26
1712	AAB65626	Homo sapiens	SUGE- Novel protein kinase, SEQ ID NO: 152.	209	82
1712	AAM25283	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:798.	209	82
1712	AAU17269	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 834.	176	67
1713	gi18256065	Mus musculus	Similar to ATPase, class II, type 9A	127	67
1713	AAM76495	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 36801.	123	70
1713	AAM63681	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 35786.	123	70
1714	gi8096269	Nicotiana tabacum	KED	149	28
1714	gi1752736	Saccharomyces cerevisiae	gene required for phosphorylation of oligosaccharides/ has high homology with YJR061w	148	30
1714	gi2292986	Rattus norvegicus	cyclic nucleotide-gated channel beta subunit	141	28
1715	AAM72995	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 33301.	158	47
1715	AAM60359	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 32464.	158	47
1715	gi 13539605 emb CAC35	Paramecium tetraurelia	cyclophilin-RNA interacting protein	144	45

183

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	733.1]				
1716	AAM71015	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 31321.	251	64
1716	AAM58517	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 30622.	251	64
1716	AAU19766	Homo sapiens	HUMA- Human novel extracellular matrix protein, Seq ID No 416.	161	44
1718	gi1420924	Zea mays	IN1	75	27
1718	gi14521970  refNP_1274 47.1	Pyrococcus abyssi	O-sialoglycoprotein endopeptidase	73	35
1719	gi20513851	Hordeum vulgare	BPM	74	35
1719	gi21039126	Cryptosporidium parvum	60 kDa glycoprotein	74	26
1719	gi207158	Rattus norvegicus	big tau	73	36
1720	gi18181943	Caenorhabditis elegans	heparan sulfate GlcNAc transferase-I/II	67	34
1720	gi2058699	Caenorhabditis elegans	multiple exostoses homolog 2	67	34
1720	gi17554740  refNP_4993 68.1	Caenorhabditis elegans	MULTIPLE EXOSTOSES HOMOLOG 2	67	34
1721	AAM69150	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29456.	200	38
1721	AAM56769	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 28874.	200	38
1721	gi4185947	Human endogenous retrovirus K	pol protein	196	38
1722	gi2065210	Mus musculus	Pro-Pol-dUTPase polyprotein	615	60
1722	gi18676710	Homo sapiens	FLJ00254 protein	592	60
1722	gi20469453  refXP_1140 40.1	Homo sapiens	similar to FLJ00254 protein	283	50
1723	gi13881755	Mycobacterium tuberculosis CDC1551	cation efflux system protein	74	30
1724	AAG78866	Homo sapiens	SHAN- Human zinc finger protein 15.	141	68
1724	ABB17928	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 6585.	99	53
1724	gi21295712  gb EAA078 57.1	Anopheles gambiae str. PEST	agCP1631	75	26
1725	gi21104340	Homo sapiens	obscurin	1586	83
1725	gi7024535	Gallus gallus	structural muscle protein titin	207	24
1725	gi1513030	Gallus gallus	connectin/titin	207	24
1727	AAE19162	Homo sapiens	THOR/ Human kinase polypeptide (PKIN-20).	1096	99

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1727	gi2736151	Rattus norvegicus	mytonic dystrophy kinase-related Cdc42-binding kinase	902	78
1727	gi1695873	Homo sapiens	ser-thr protein kinase PK428	896	77
1728	AAAY99411	Homo sapiens	GETH Human PRO1487 (UNQ756) amino acid sequence SEQ ID NO:260.	862	67
1728	gi15617453	Homo sapiens	chondroitin synthase	862	67
1728	AAE15959	Homo sapiens	EUMO- Human 4589624/92-303 protein, member of Fringe and Brainiac family.	761	79
1729	gi 15804980 ref NP_290960.1	Escherichia coli O157:H7 EDL933	Uncharacterized conserved protein	71	33
1731	gi14268490	Musca domestica	hunchback	82	33
1731	AAM93401	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3002.	76	27
1731	gi2076606	Musca domestica	hunchback zinc finger protein	73	30
1732	AAAY91949	Homo sapiens	INCY- Human cytoskeleton associated protein 4 (CYSKP-4).	1047	57
1732	ABB90754	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 240.	1043	57
1732	gi619577	Gallus gallus	cardiac muscle tensin	1043	56
1733	gi3090889	Homo sapiens	synapsin IIIa	70	38
1733	gi6572355	Homo sapiens	cE86D10.1 (synapsin III)	70	38
1733	gi 19924105 ref NP_003481.2	Homo sapiens	synapsin III, isoform IIIa	70	38
1734	AAB85144	Homo sapiens	HUMA- Human NKCR polypeptide (clone ID HMSOM53).	1506	93
1734	gi4973126	Mus musculus castaneus	high affinity immunoglobulin gamma Fc receptor I	490	39
1734	gi4973124	Mus musculus	high affinity immunoglobulin gamma Fc receptor I	489	39
1735	gi 15597595 ref NP_251089.1	Pseudomonas aeruginosa	pyoverdine synthetase D	69	30
1736	gi14488302	Oryza sativa	Putative transposon protein	81	24
1736	gi3851516	Phytophthora infestans	cyst germination specific acidic repeat protein precursor	72	33
1736	gi 14488302 gb AAK63883.1 AC074105.12	Oryza sativa	Putative transposon protein	81	24
1737	AAB85357	Homo sapiens	INCY- Human phosphatase (PP) (clone ID 3402521CD1).	1591	100
1737	gi21205864	Homo sapiens	T-cell activation protein phosphatase 2C; TA-PP2C	1591	100
1737	gi21464366	Drosophila melanogaster	RE06653p	758	52
1738	gi7271811	Drosophila melanogaster	GTPase activating protein	292	38
1738	AAM76430	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 36736.	246	100
1738	AAM63615	Homo sapiens	MOLE- Human brain expressed single	246	100

185  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			exon probe encoded protein SEQ ID NO: 35720.		
1739	ABB50365	Homo sapiens	HUMA- Human secreted protein encoded by gene 65 SEQ ID NO:313.	272	87
1739	AAW88598	Homo sapiens	HUMA- Secreted protein encoded by gene 65 clone HFVHY45.	272	87
1739	ABB50764	Homo sapiens	HUMA- Human secreted protein encoded by gene 65 SEQ ID NO:716.	143	92
1740	gi2065210	Mus musculus	Pro-Pol-dUTPase polyprotein	1210	58
1740	gi10834720 gb AAG23790.1 AF258587.1	Homo sapiens	PP565	274	80
1740	gi385615 gb AAB26708.1	Mus sp.	fibulin gene homolog	248	75
1741	ABB90748	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 228.	2116	97
1741	gi15987493	Homo sapiens	tumor endothelial marker 6	2116	97
1741	ABB90754	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 240.	530	37
1742	ABB11753	Homo sapiens	HYSE- Human NOV/plexin-A1 homologue, SEQ ID NO:2123.	291	90
1742	gi1665757	Mus musculus	plexin 1	291	90
1742	gi6010217	Homo sapiens	NOV/plexin-A1 protein	291	90
1743	AAM79514	Homo sapiens	HYSE- Human protein SEQ ID NO 3160.	149	90
1743	AAM78530	Homo sapiens	HYSE- Human protein SEQ ID NO 1192.	149	90
1743	gi1244510	Homo sapiens	p311 protein	149	90
1744	AAG93324	Homo sapiens	NISC- Human protein HP10370.	83	41
1744	gi21064771	Drosophila melanogaster	RH61467p	83	46
1744	gi18676554	Homo sapiens	FLJ00174 protein	77	41
1745	gi4128039	Homo sapiens	TL132 protein	81	29
1745	gi17983118	Brucella melitensis	METAL DEPENDENT HYDROLASE	74	23
1745	AAU75578	Homo sapiens	UYNA- Human ubiquitin specific protease 10 (USP10).	71	31
1746	gi15074154	Sinorhizobium meliloti	PUTATIVE FATTY ACID/PHOSPHOLIPID SYNTHESIS PROTEIN	76	25
1746	gi1869833	human herpesvirus 2	myristylated tegument protein	75	27
1746	gi20516045	Thermoanaerobacter tengcongensis	Chemotaxis response regulator CheB, consists of CheY-like receiver domain and a methyltransferase (demethylase) domain	69	20
1747	gi18025496	cercopithecine herpesvirus 15	EBNA-1	124	37
1747	gi5821153	Homo sapiens	RNA binding protein	123	29
1747	gi6649242	Homo sapiens	splicing coactivator subunit SRm300	123	29
1748	gi4321764 gb AAD1581	Mus musculus	MAP kinase kinase 7 alpha 2	65	30



186  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	9.1]				
1748	gi 20859704 ref XP_133986.1	Mus musculus	mitogen activated protein kinase kinase 7	65	30
1748	gi 4321768 gb AAD1582.1.1	Mus musculus	MAP kinase kinase 7 beta 2	65	30
1749	AAB50964	Homo sapiens	GETH Human PRO1313 protein.	439	89
1749	AAB47290	Homo sapiens	GETH PRO1313 polypeptide.	439	89
1749	AAB24431	Homo sapiens	GETH Human PRO1313 protein sequence SEQ ID NO:216.	439	89
1750	AAU00502	Homo sapiens	MILL- Human TANGO 437 protein.	115	91
1750	gi 20384654	Homo sapiens	two-pore calcium channel protein 2	115	91
1750	AAM91059	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:18652.	93	64
1751	gi 10440494	Homo sapiens	FLJ00092 protein	252	97
1751	AAM40956	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5887.	80	30
1751	gi 10440494 dbj BAB15780.1	Homo sapiens	FLJ00092 protein	252	97
1752	gi15980036	Yersinia pestis	2-dehydro-3-deoxyphosphooctonate aldolase	77	46
1752	gi11322261	Diceros bicornis	alpha adrenergic receptor 2B	74	26
1752	gi20516240	Thermoanaerobacter tengcongensis	methyiaspartate mutase	73	25
1753	gi19684014	Homo sapiens	similar to brain-specific angiogenesis inhibitor 3 (H. sapiens)	1387	99
1753	AAB88367	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0101.	1380	99
1753	gi1469936	Mus musculus	FGF-binding protein	158	29
1754	AAB01397	Homo sapiens	INCY- Neuron-associated protein.	435	92
1754	gi21218140	Homo sapiens	rab effector MYRIP	435	92
1754	gi21320161	Mus musculus	exophilin 8	378	77
1755	AAM74815	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 35121.	253	75
1755	AAM62013	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 34118.	253	75
1755	AAM70390	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 30696.	228	62
1756	gi6460201	Deinococcus radiodurans	phenylacetic acid degradation protein PaaA	85	27
1756	gi3309543	Takifugu rubripes	MLL	79	34
1756	AAT10059_aal	Homo sapiens	USSH erbB-3 cDNA clone E3-16.	74	31
1757	gi18676406	Homo sapiens	FLJ00021 protein	70	36
1758	gi13423395	Caulobacter crescentus CB15	NADH dehydrogenase I, M subunit	78	37

187

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1758	gi 17506337 ref NP_491390.1	Caenorhabditis elegans	D1007.15.p	82	24
1758	gi 16126181 ref NP_420745.1	Caulobacter crescentus CB15	NADH dehydrogenase I, M subunit	78	37
1759	gi19881193	chimpanzee cytomegalovirus	transcriptional transactivator TRS1	83	29
1759	gi19881161	chimpanzee cytomegalovirus	transcriptional transactivator IRS1	83	29
1759	gi556297	Mus musculus	alpha-1 type IV collagen	81	33
1760	gi18033185	Danio rerio	UNC45-related protein	702	79
1760	AAG77802	Homo sapiens	HUMA- Human HOGEN50 serine/threonine phosphatase protein sequence.	603	65
1760	AAM40290	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3435.	603	65
1761	gi6634123	Drosophila melanogaster	SoxNeuro	70	24
1762	gi 14245700 dbj BAB56142.1	Giardia intestinalis	kinesin-like protein 4	69	26
1762	gi 165011 gb AAA31246.1	Oryctolagus cuniculus	eucaryotic release factor (eRF)	69	24
1762	gi 15559188 emb CAC03424.2	Homo sapiens	dJ45P21.3 (butyrophilin, subfamily 3, member A1)	69	26
1763	AAM93661	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3536.	186	80
1763	AAM64398	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 36503.	154	76
1763	gi 20556958 ref XP_061562.5	Homo sapiens	similar to PAM COOH-terminal interactor protein 1	73	43
1764	AAU17223	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 788.	211	87
1765	gi1334546	Podospora anserina	Dod COI i13 grp IB protein	71	37
1765	gi5679307	Mus musculus	RORgamma t	70	27
1765	gi4186077	Mus musculus	ROR gamma T protein	70	27
1766	gi17864081	Mus musculus	PPAR gamma coactivator-1beta protein	74	26
1766	gi44795	Methanococcus voltae	polyferredoxin	71	28
1766	gi14279670	Lycopersicon esculentum	verticillium wilt disease resistance protein	71	31
1768	AAE06588	Homo sapiens	SAGA Human protein having hydrophobic domain, HP10778.	165	100
1768	AAM40979	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5910.	165	100
1768	AAB24542	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 27 SEQ ID NO:168.	73	30

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1769	gi6174840	Achromobacter xylosoxidans subsp. xylosoxidans	low-specificity D-threonine aldolase	78	33
1769	gi16769806	Drosophila melanogaster	SD02660p	75	23
1769	gi1098473	Rattus norvegicus	insulin-like growth factor binding protein	73	31
1770	AAP94684	Homo sapiens	CHIL Amino acid sequence encoded by part of human mannose binding protein(hMBP) genomic DNA.	79	56
1770	gi 15790548 ref NP_280372.1	Halobacterium sp. NRC-1	cobryic acid synthase; CbiP	69	36
1770	gi 11467609 ref NP_050661.1	Guillardia theta	Clp protease ATP binding subunit	69	27
1772	gi5532460	Shigella flexneri	ShiF	66	32
1773	gi11544663	Arabidopsis thaliana	PTPKIS1	75	42
1773	gi11595504	Arabidopsis thaliana	PTPKIS1 protein	75	42
1773	gi18389331	Mus musculus	2',5'-oligoadenylate synthetase-like 10	73	42
1774	AAM06519	Homo sapiens	HYSE- Human foetal protein, SEQ ID NO: 250.	414	90
1774	gi 18552248 ref XP_092510.1	Homo sapiens	similar to latent transforming growth factor beta binding protein 1; latent TGF beta binding protein	69	37
1775	gi4884924	Rangiferine herpesvirus 1	glycoprotein C	67	60
1775	AAB94152	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14435.	65	34
1775	AAB93253	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12271.	65	34
1776	gi13424176	Caulobacter crescentus CB15	N-carbamyl-L-amino acid amidohydrolase	89	24
1776	gi514267	Homo sapiens	proto-oncogene tyrosine-protein kinase	86	29
1776	gi28237	Homo sapiens	p150 protein (AA 1-1130)	84	28
1777	gi63370	Gallus gallus	dystrophin (AA 1 - 3660)	68	31
1777	gi 3046783 emb CAA68033.1	Scyliorhinus canicula	dystrophin	67	29
1777	gi 2342682 gb AAB70406.1	Arabidopsis thaliana	Contains similarity to Rattus AMP-activated protein kinase (gb X95577).	67	31
1778	AAE16176	Homo sapiens	INCY- Human G-protein coupled receptor 7 (GCREC-7) protein.	1419	100
1778	AAE18021	Homo sapiens	CURA- Human G-protein coupled receptor-8a (GPCR-8a) protein.	1419	100
1778	AAG72411	Homo sapiens	YEDA Human OR-like polypeptide query sequence, SEQ ID NO: 2092.	1419	100
1779	AAM76040	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 36346.	93	48

189

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1779	AAM63227	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 35332.	93	48
1779	gi12620576	Bradyrhizobium japonicum	ID342	87	24
1780	gi2459833	Rattus norvegicus	Maxp1	81	31
1780	AAB65650	Homo sapiens	SUGE- Novel protein kinase, SEQ ID NO: 177.	80	35
1780	AAM39805	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2950.	80	36
1781	gi4877963	Mus musculus	NF-kappaB inducing kinase	69	39
1781	gi15077865	Mus musculus	bullous pemphigoid antigen 1-b	67	35
1781	gi15077863	Mus musculus	bullous pemphigoid antigen 1-a	67	35
1782	gi4138265	Nicotiana tabacum	Avr9 elicitor response protein	76	27
1782	gi12725153	Lactococcus lactis subsp. lactis	50S ribosomal protein L3	75	32
1782	AAB21008	Homo sapiens	INCY- Human nucleic acid-binding protein, NuABP-12.	73	32
1783	gi3947714	Streptococcus agalactiae	initiation factor IF2	86	20
1783	gi9558387	Streptococcus agalactiae	initiation factor 2	86	20
1783	gi9558369	Streptococcus agalactiae	initiation Factor 2	86	20
1786	gi435855	Mus sp.	CREB-binding protein; CBP	75	22
1786	gi2911464	Leishmania tarentolae	sodium stibogluconate resistance protein	75	34
1786	gi19547887	Mus musculus	CREB-binding protein	75	22
1787	gi3747099	Mus musculus	C1q-related factor	616	61
1787	gi14278927	Mus musculus	gliacolin	615	64
1787	gi10566471	Mus musculus	Gliacolin	615	64
1788	gi 21291197 gb EAA033.42.1	Anopheles gambiae str. PEST	agCP7579	71	20
1788	gi 20803964 emb CAD31541.1	Mesorhizobium loti	HYPOTHETICAL PROTEIN	69	43
1789	AAM41125	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6056.	320	80
1789	AAM39339	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2484.	320	80
1789	AAM79857	Homo sapiens	HYSE- Human protein SEQ ID NO 3503.	320	80
1790	gi1143585	Paracentrotus lividus	2 alpha fibrillar collagen	69	23
1791	gi9837427	Lytechinus variegatus	embryonic blastocoelar extracellular matrix protein precursor	116	34
1791	gi14089698	Mycoplasma pulmonis	OLIGOPEPTIDE ABC TRANSPORTER PERMEASE PROTEIN	71	23
1791	gi6572111	Bartonella	riboflavin synthase alpha chain	69	29

190

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		quintana			
1792	gi 4506023 ref NP_002710.1	Homo sapiens	protein phosphatase 2, regulatory subunit B (B56), gamma isoform	68	39
1793	AAM711170	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 31476.	180	82
1793	AAM58664	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 30769.	180	82
1793	AAM65679	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 37784.	168	71
1794	AAG00072	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4153.	125	80
1794	AAW34618	Homo sapiens	IMUT- Human C3 protein mutant DV-7N.	125	80
1794	AAW34617	Homo sapiens	IMUT- Human C3 protein mutant DV-6.	125	80
1795	AAV05069	Homo sapiens	SMIK Human PIGR-2 protein sequence.	1055	85
1795	gi396170	Homo sapiens	CMRF-35 antigen	406	45
1795	gi18490143	Homo sapiens	CMRF35 leukocyte immunoglobulin-like receptor	406	45
1796	gi 6723273 dbj BAA89659.1	Baboon endogenous virus strain M7	gag-pol precursor polyprotein	421	41
1796	gi 13940448 gb AAK50381.1 U43202.2	Murine leukemia virus	pol precursor protein	421	41
1796	gi 331995 gb AAB03091.1	AKV murine leukemia virus	gag-pol polyprotein (tag amber codon at 2250-2252 inserts Gln in Mo-MuLV)	421	41
1797	gi21411325	Homo sapiens	Similar to LOC205103	260	73
1797	gi 4835878 gb AAD30280.1 AF134838.1	Homo sapiens	endocytic receptor Endo180	77	31
1797	gi 16076075 emb CAC94295.1	Leishmania donovani donovani	trypanothione reductase	70	30
1798	gi927721	Saccharomyces cerevisiae	Sip1p: SNF1 protein kinase substrate; YDR422C; CAI: 0.13	72	34
1798	gi172604	Saccharomyces cerevisiae	protein kinase	72	34
1798	gi 6320630 ref NP_010710.1	Saccharomyces cerevisiae	SNF1 protein kinase substrate; Sip1p	72	34
1799	gi 20839768 ref XP_130311.1	Mus musculus	similar to GDP-fucose transporter 1	71	29
1801	gi 17461642 ref XP_0662	Homo sapiens	similar to Ig kappa chain	78	23

191  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	49.1				
1801	gi 6325342 ref NP_015410.1	Saccharomyces cerevisiae	Protein required for cell viability; Ypr085cp	76	22
1801	gi 9635081 ref NP_057809.1	Gallid herpesvirus 2	UL47	74	26
1802	AAB94148	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14427.	250	56
1802	AAG64564	Homo sapiens	SHAN- Human zinc-finger protein 60.	250	56
1802	AAM79356	Homo sapiens	HYSE- Human protein SEQ ID NO 3002.	250	56
1803	AAW81754	Homo sapiens	BOEF Human Fanconi anaemia-associated gene II protein.	631	85
1803	gi2407911	Homo sapiens	differentially expressed in Fanconi anemia	555	74
1803	gi6013073	Mus musculus	HemT-3 protein	89	24
1805	gi14189735	Homo sapiens	ATP-binding cassette transporter family A member 12	1508	90
1805	gi1943947	Bos taurus	ABC transporter	404	31
1805	AAZ94734_aa1	Homo sapiens	FARB Human ATP binding cassette ABCA1 (ABC1) cDNA.	395	33
1806	AAU12234	Homo sapiens	GETH Human PRO4350 polypeptide sequence.	859	100
1806	AAA96344_aa1	Homo sapiens	GETH cDNA encoding a novel polypeptide designated PRO4357.	498	48
1806	AAU12445	Homo sapiens	GETH Human PRO4357 polypeptide sequence.	498	48
1807	gi190396	Homo sapiens	profilaggrin	76	29
1808	AAB88367	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0101.	74	30
1808	gi19684014	Homo sapiens	similar to brain-specific angiogenesis inhibitor 3 (H. sapiens)	74	30
1808	gi 18576362 ref XP_084481.1	Homo sapiens	similar to fibroblast growth factor binding protein 1	74	30
1809	gi530876	Chlamydomonas reinhardtii	amino acid feature: Rod protein domain, aa 266 .. 468; amino acid feature: globular protein domain, aa 32 .. 265	126	35
1809	gi6578849	Myxococcus xanthus	FrgA	126	29
1809	gi2429362	Santalum album	proline rich protein	122	27
1810	gi17428288	Ralstonia solanacearum	PROBABLE CATION-TRANSPORTING ATPASE LIPOPROTEIN TRANSMEMBRANE	75	28
1810	gi21483422	Drosophila melanogaster	LD34142p	71	29
1810	ABB90042	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2418.	70	32
1811	gi 20915248 ref XP_145160.1	Mus musculus	similar to Collagen alpha 1(VI) chain precursor	148	74
1812	gi2104558	Rattus	CCA3	1150	90

192

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		norvegicus			
1812	AAB64963	Homo sapiens	ROSE/ Human secreted protein sequence encoded by gene 24 SEQ ID NO:141.	172	37
1812	gi12963869	Mus musculus	gene trap ankyrin repeat containing protein	172	37
1813	AAB65201	Homo sapiens	GETH Human PRO1009 (UNQ493) protein sequence SEQ ID NO:194.	208	100
1813	AAV66678	Homo sapiens	GETH Membrane-bound protein PRO1009.	208	100
1813	AAB24068	Homo sapiens	GETH Human PRO1009 protein sequence SEQ ID NO:36.	208	100
1815	AAG89314	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 434.	191	100
1815	gi6460052	Deinococcus radiodurans	dipeptidyl peptidase IV-related protein	66	60
1816	gi1052594	Drosophila melanogaster	trithorax protein trxl	75	26
1816	gi1052593	Drosophila melanogaster	trithorax protein trxII	75	26
1816	gi158818	Drosophila melanogaster	zinc-binding protein	75	26
1817	AAB49765	Homo sapiens	HELI- Human proliferation differentiation factor amino acid sequence.	229	94
1817	AAB88393	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0137.	229	94
1817	gi18446895	Drosophila melanogaster	AT05866p	73	25
1818	gi6573212	Giardia intestinalis	variant-specific surface protein H7-1	73	32
1818	gi159143	Giardia intestinalis	variant-specific surface protein H7	73	32
1818	gi15144254	Micrurus corallinus	neurotoxin homologue 8	72	32
1819	gi161857	Tetrahymena thermophila	surface antigen	69	35
1821	gi913964	Carcinoscorpius rotundicauda	factor C	80	26
1821	gi217397	Tachypleus tridentatus	limulus factor C precursor	80	26
1821	gi18542425	Tachypleus tridentatus	factor C precursor	80	26
1822	gi9309473	Mus musculus	DNMT1 associated protein-1	74	37
1822	gi1666895	Homo sapiens	CHL1 protein	74	23
1822	gi16923930	Mus musculus	MAT1-mediated transcriptional repressor	74	37
1823	gi9058659	Canis familiaris	skeletal muscle chloride channel ClC-1	73	34
1823	gi433182	Drosophila melanogaster	receptor protein tyrosine phosphatase	72	26
1823	gi20429105	Paracoccus zeaxanthinifaciens	decaprenyl diphosphate synthase	72	27
1824	gi13374178	Mus musculus	TAFII140 protein	612	88

193

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1824	gi17861888	Drosophila melanogaster	GM10839p	246	49
1824	gi6634096	Drosophila melanogaster	BIP2 protein	242	48
1825	gi16605480	Homo sapiens	G6b-C protein	1159	100
1825	gi16605484	Homo sapiens	G6b-E protein	1009	90
1825	gi5304877	Homo sapiens	immunoglobulin receptor	1003	83
1826	AAB94636	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15515.	105	37
1826	AAU15903	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 856.	105	37
1826	gi21430928	Drosophila melanogaster	SD27341p	93	39
1827	AAR33270	Homo sapiens	WIST- T cell receptor alpha chain clone alpha1.3.	329	92
1827	gi1806100	Homo sapiens	T cell receptor alpha chain	329	92
1827	gi2358032	Homo sapiens	TCRAV8S3	329	92
1828	gi20513851	Hordeum vulgare	BPM	73	45
1828	AAO01897	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15789.	70	35
1828	AAE16477	Homo sapiens	OSTE- Human collagen alpha1 (II) protein.	69	31
1829	AAG66837	Homo sapiens	SHAN- Human ATP-dependent serine proteinase 31.	356	100
1829	AAG66838	Homo sapiens	SHAN- Human ATP-dependent serine proteinase 31 N-terminal peptide.	89	100
1829	gi5881591	Gallus gallus	homeodomain protein	77	38
1830	AAB94294	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14745.	951	99
1830	gi10504968	Drosophila melanogaster	rho guanine nucleotide exchange factor 4	180	22
1830	gi16197921	Drosophila melanogaster	LD03170p	180	22
1831	ABB12353	Homo sapiens	HYSE- Human bone marrow expressed protein SEQ ID NO: 107.	199	30
1831	gi20452161	Canis familiaris	retinitis pigmentosa GTPase regulator	143	24
1831	gi2062609	Xenopus laevis	middle molecular weight neurofilament protein NF-M(1)	140	24
1832	AAB29778	Homo sapiens	RHOD- Human MSF-derived tribonectin.	148	18
1832	gi142161	Anaplasma marginale	surface antigen Amf105	141	25
1832	gi4808177	Drosophila subobscura	largest subunit of the RNA polymerase II complex	141	20
1833	AAM66321	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26627.	424	51
1833	AAM53933	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26038.	424	51
1833	gi16723273[d bj]BAA8965 9.1]	Baboon endogenous virus strain M7	gag-pol precursor polypeptide	357	47



194

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1834	AAM88756	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:16349.	208	100
1834	gi20417	Persea americana	cellulase	77	34
1834	gi153337	Streptomyces tenebrarius	kanamycin-apramycin resistance methylase	69	26
1837	AAAY02893	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 92.	76	41
1837	AAAY99429	Homo sapiens	GETH Human PRO1563 (UNQ769) amino acid sequence SEQ ID NO:317.	73	35
1837	gi6634084	Drosophila melanogaster	malate dehydrogenase (NADP-dependent oxaloacetate decarboxylating), malic enzyme	73	39
1838	gi2865602	Saccharopolyspora sp.	SapI M2 methyltransferase	77	37
1838	gi3089358	Rattus norvegicus	MARRLC2A	75	33
1838	gi2865602 gb AAC9718.2.1	Saccharopolyspora sp.	SapI M2 methyltransferase	77	37
1839	AAM69149	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29455.	154	96
1839	AAM56768	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 28873.	154	96
1839	AAW96209	Homo sapiens	SMIK Amyloid precursor protein (APP) C-terminal fragment.	102	78
1840	gi9946563	Pseudomonas aeruginosa	probable type II secretion system protein	81	36
1840	gi21108565	Xanthomonas axonopodis pv. citri str. 306	pseudouridylylase synthase	75	35
1840	ABB04714	Homo sapiens	SHAN- Human PP1744 protein SEQ ID NO:23.	74	31
1841	gi1491949	Molluscum contagiosum virus subtype 1	MC006L	85	30
1841	AAM42085	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 7016.	81	27
1841	AAM40299	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3444.	81	27
1842	gi20381413	Homo sapiens	Similar to LOC160680	216	44
1842	gi13592175	Leishmania major	ppg3	144	24
1842	gi5420387	Leishmania major	proteophosphoglycan	140	23
1843	AAB87181	Homo sapiens	MILL- Human secreted protein MANGO 349 E41D variant, SEQ ID NO:231.	278	42
1843	AAB87128	Homo sapiens	MILL- Human secreted protein MANGO 349, SEQ ID NO:130.	278	42
1843	AAB87179	Homo sapiens	MILL- Human secreted protein MANGO 349 I21K variant, SEQ ID	276	41

195

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO:227.		
1844	AAE14341	Homo sapiens	INCY- Human protease PRPS-6 protein.	886	93
1844	gi16768276	Drosophila melanogaster	GH27809p	290	41
1844	gi2655204	Mus musculus	ubiquitin-specific protease	258	35
1846	AAAY88300	Homo sapiens	MILL- Human TANGO 187-3 protein.	1334	90
1846	gi13097780	Homo sapiens	Similar to RIKEN cDNA 2810037C14 gene	1326	90
1846	AAAY88296	Homo sapiens	MILL- Human TANGO 187-2/3 protein.	1312	87
1847	AAG74984	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5748.	75	32
1847	gi17352449	Rattus norvegicus	ErbB3/Her3 precursor	74	38
1847	gi 20860870 ref XP_125664.1	Mus musculus	similar to H4(D10S170) protein	75	32
1848	gi3123530	Fowlpox virus	fpI3L, orthologue of vaccinia I3L	75	27
1848	gi5902659	Drosophila melanogaster	ring canal protein	70	27
1848	gi 18110218 ref NP_476589.2	Drosophila melanogaster	kel-P2	70	27
1849	gi2065210	Mus musculus	Pro-Pol-dUTPase polypeptide	614	78
1849	AAM65715	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26021.	548	73
1849	AAM53338	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25443.	548	73
1850	gi10999071	Lophognathus longirostris	NADH dehydrogenase subunit 2	74	23
1850	gi18537243	Human immunodeficiency virus type 1	envelope glycoprotein	74	29
1850	gi 10999071 gb AAG00622.2 AF128462.2	Lophognathus longirostris	NADH dehydrogenase subunit 2	74	23
1851	gi 17448210 ref XP_068503.1	Homo sapiens	similar to 60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat shock protein 60) (HSP-60) (Mitochondrial matrix protein P1) (P60 lymphocyte protein) (HuCHA60)	72	28
1852	gi1164937	Saccharomyces cerevisiae	YOR3160w	74	31
1852	gi3176662	Arabidopsis thaliana	Similar to mannosyl-oligosaccharide glucosidase gb X87237 from Homo sapiens.	73	31
1852	gi13398928	Arabidopsis thaliana	alpha-glucosidase 1	73	31
1853	gi 20889364	Mus musculus	similar to hepatitis A virus cellular	76	36

196

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	refXP_138429.1		receptor 1; T cell immunoglobulin domain and mucin doamin protein 1		
1853	gi 21288202 gb EAA00523.1	Anopheles gambiae str. PEST	agCP9342	71	32
1854	AAB88481	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0251.	776	99
1854	AAE03835	Homo sapiens	HUMA- Human gene 18 encoded secreted protein HFKHW50, SEQ ID NO: 81.	776	99
1854	AAE03863	Homo sapiens	HUMA- Human gene 18 encoded secreted protein HFKHW50, SEQ ID NO:109.	716	97
1855	gi1663748	Chlamydomonas reinhardtii	dynein heavy chain 7	82	29
1855	gi1663744	Chlamydomonas reinhardtii	dynein heavy chain 5	80	28
1855	gi1663738	Chlamydomonas reinhardtii	dynein heavy chain 2	80	27
1856	gi18032120	Gallus gallus	shal-like voltage-gated potassium channel	75	23
1856	gi1408569	Haemophilus influenzae	adhesion and penetration protein	71	28
1856	gi 18032120 gb AAL56633.1 AF075160.1	Gallus gallus	shal-like voltage-gated potassium channel	75	23
1857	AAM67180	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27486.	129	44
1857	AAM54795	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26900.	129	44
1857	gi 21040255 ref NP_631907.1	Homo sapiens	splicing factor, arginine/serine-rich 12	109	29
1858	gi21392190	Drosophila melanogaster	RE74758p	71	39
1858	gi9954108	Trypanosoma cruzi	RNA binding protein RGGm	68	40
1858	gi20302994	Medicago truncatula	nodule-specific glycine-rich protein 1C	66	32
1859	gi 20536244 ref XP_060505.4	Homo sapiens	similar to autoantigen La	72	30
1860	gi 17541362 ref NP_502409.1	Caenorhabditis elegans	K08E7.5.p	103	29
1860	gi 17446900 ref XP_065833.1	Homo sapiens	similar to DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - Mastigamoeba invertens (fragment)	100	34
1860	gi 9628166 ref NP_0427	African swine fever virus	CD2 homolog	98	30

197  
Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	52.1				
1861	AAAY70691	Homo sapiens	DAND Human membrane attractin-2.	162	40
1861	AAAY70690	Homo sapiens	DAND Human membrane attractin-1.	162	40
1861	gi12275390	Rattus norvegicus	membrane attractin	162	40
1862	gi10039425	Equus caballus	ALR protein	81	28
1862	gi13529521	Mus musculus	Similar to elastin microfibril interface located protein	80	32
1862	AAM40414	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3559.	79	39
1863	gi 16588389 gb AAL26787.1 AF304442.1	Homo sapiens	B lymphocyte activation-related protein BC-1514	247	52
1863	gi 20479028 ref XP_113729.1	Homo sapiens	similar to B lymphocyte activation-related protein BC-1514	117	68
1863	gi 21301715 gb EAA13860.1	Anopheles gambiae str. PEST	agCP8366	85	41
1864	AAU15851	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 804.	1275	78
1864	AAU16312	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1265.	1123	76
1864	AAG02054	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6135.	308	91
1865	AAB94953	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16485.	86	29
1865	gi3746787	Homo sapiens	SYT interacting protein SIP	86	29
1865	gi15022507	Homo sapiens	coactivator activator	86	29
1866	gi17133332	Nostoc sp. PCC 7120	preprotein translocase SecY subunit	68	43
1866	gi 13489110 ref NP_068773.1	Homo sapiens	gap junction protein, alpha 3, 46kD (connexin 46)	66	40
1867	gi706930	Rattus norvegicus	cyclic GMP stimulated phosphodiesterase	191	95
1867	AAV54762_aal	Homo sapiens	UNIW Human cGS-PDE cDNA DNA sequence.	137	100
1867	AAV36157_aal	Homo sapiens	UNIW Human cyclic-GMP-nucleotide phosphodiesterase cDNA.	137	100
1868	AAB95695	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18516.	112	27
1868	AAV91447	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 48 SEQ ID NO:168.	112	27
1868	AAV91393	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 48 SEQ ID NO:114.	112	27
1870	AAU07886	Homo sapiens	WHED Polypeptide sequence for human hspG15.	1454	94
1870	gi13603891	Homo sapiens	MOV10-like 1	1454	94
1870	gi13603857	Mus musculus	MOV10-like 1	954	77
1871	AAM96652	Homo sapiens	HUMA- Human reproductive system	484	96

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			related antigen SEQ ID NO: 5310.		
1871	gi18676652	Homo sapiens	FLJ00225 protein	433	95
1871	gi21386760	Berneuxia tibetica	maturase R	70	32
1872	AAQ90304_aal	Homo sapiens	NISR Human thyroid peroxidase gene.	73	29
1872	AAW48781	Homo sapiens	RSRR- Thyroid peroxidase.	73	29
1872	AAR75689	Homo sapiens	NISR Human thyroid peroxidase.	73	29
1873	AAG03774	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7855.	228	90
1873	gi338288	Homo sapiens	preprosomatostatin I	228	90
1873	gi342299	Macaca fascicularis	preprosomatostatin	228	90
1875	AAR30418	Homo sapiens	DAND Nearly complete p107 protein.	76	30
1875	gi347378	Homo sapiens	p107	76	30
1875	gi157871	Drosophila melanogaster	P glycoprotein	76	24
1876	ABB17955	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 6612.	186	40
1876	AAS17764_aal	Homo sapiens	GENA- Human Genomic DNA for CRYBB1.	167	39
1876	AAO02331	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 16223.	165	42
1877	gi59977 emb CAA7866 2.1	Human endogenous retrovirus	tripartite fusion transcript PLA2L	224	76
1878	ABB84943	Homo sapiens	GETH Human PRO1556 protein sequence SEQ ID NO:254.	1056	93
1878	AAB31670	Homo sapiens	PROT- Amino acid sequence of a human protein having a hydrophobic domain.	1056	93
1878	AAB47295	Homo sapiens	GETH PRO1556 polypeptide.	1056	93
1879	ABB15861	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 4518.	73	36
1880	AAU83117	Homo sapiens	ZYMO Novel secreted protein Z799543G2P.	66	54
1880	gi12723186	Lactococcus lactis subsp. lactis	outer membrane lipoprotein precursor	66	26
1881	gi609624	Vibrio cholerae	EpsC	73	29
1882	gi12667456	Rattus norvegicus	synaptotagmin VIIId	86	32
1882	gi12667454	Rattus norvegicus	synaptotagmin VIIc	85	33
1882	gi334072	Pseudorabies virus	ORF-3 protein	83	35
1883	gi1747	Oryctolagus cuniculus	trichohyalin	119	29
1883	gi2072290	Xenopus laevis	XL-INCENP	100	27
1883	gi12584554	Human coxsackievirus B3	polyprotein	96	25
1884	gi15601413 refNP 2330	Vibrio cholerae	sucrose-6-phosphate dehydrogenase	65	55

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	44.1]				
1885	gi16878287	Homo sapiens	Similar to C-terminal modulator protein	74	35
1885	gi15866714	Homo sapiens	C-terminal modulator protein	74	35
1885	AAO06984	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 20876.	70	60
1887	AAW25939	Homo sapiens	CNRS T-cell receptor V-beta-5.1 peptide fragment.	601	99
1887.	gi36973	Homo sapiens	T-cell receptor beta-chain	601	99
1887	gi1552498	Homo sapiens	V segment translation product	600	100
1888	gi18874468	Homo sapiens	partitioning-defective 3-like protein splice variant c	198	73
1888	gi16903870	Homo sapiens	partitioning-defective 3-like protein splice variant b	198	73
1888	gi16903868	Homo sapiens	partitioning-defective 3-like protein splice variant a	198	73
1889	gi21489377	Homo sapiens	MAPA protein	1620	99
1889	gi21489330	Bos taurus	MAPA protein	833	56
1889	gi21489379	Mus musculus	MAPA protein	630	48
1890	AAY10874	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	503	100
1890	gi17429674	Ralstonia solanacearum	PROBABLE LIPOPROTEIN	73	44
1891	gi15723141	Homo sapiens	c349E10.1.1 (novel protein, isoform 1)	180	46
1891	AAB59006	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 714.	174	47
1891	gi19353342	Mus musculus	RIKEN cDNA 9530058B02 gene	162	47
1892	AAM86086	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:13679.	95	53
1892	AAO05973	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 19865.	94	82
1892	AAO09418	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 23310.	91	70
1893	gi8778607	Arabidopsis thaliana	F5M15.23	71	25
1894	AAM65951	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26257.	69	38
1894	AAM53568	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25673.	69	38
1894	gi 20832567 ref XP_133524.1	Mus musculus	similar to Heterogeneous nuclear ribonucleoprotein A3 (hnRNP A3) (D10S102)	163	76
1895	AAM66299	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26605.	440	83
1895	AAM53913	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26018.	440	83
1895	gi 6723273 dbj BAA89659.1	Baboon endogenous virus strain M7	gag-pol precursor polypeptide	270	45

200

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1896	gi4883988	Bartonella clarridgeiae	cell division protein FtsZ	68	28
1897	AAO13209	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 27101.	142	54
1897	AAM66708	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27014.	124	46
1897	AAM54310	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26415.	124	46
1898	gi2565268	Drosophila virilis	pore-forming protein MIP family	75	27
1898	gi7453547	Homo sapiens	glioma tumor suppressor candidate region protein 1	75	31
1898	gi3218331	Metarhizium anisopliae	nitrogen response regulator	74	26
1899	gi9656609	Vibrio cholerae	chemotaxis protein CheA	73	32
1899	gi 20908537 ref XP_127414.1	Mus musculus	RIKEN cDNA 1700001L19	443	80
1899	gi 15642063 ref NP_231695.1	Vibrio cholerae	chemotaxis protein CheA	73	32
1900	gi 18586105 ref XP_091400.1	Homo sapiens	similar to sca1	203	84
1900	gi 20888279 ref XP_146508.1	Mus musculus	similar to spinocerebellar ataxia type 1	199	82
1901	gi338033	Homo sapiens	serum protein	90	32
1901	gi4808221	Homo sapiens	dJ1177I5.2 (serum constituent protein MSE55)	90	32
1901	gi4098993	Mus musculus	polyhomeotic 2	88	30
1902	AAB19933	Homo sapiens	INCY- Human oxidoreductase OXRD-8.	250	100
1902	gi19713043	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	Iron/zinc/copper-binding protein	73	22
1902	gi 20342079 ref XP_110614.1	Mus musculus	RIKEN cDNA 1700003E16	77	25
1903	gi342279	Macaca nemestrina	opiomelanocortin	231	49
1903	gi28342	Homo sapiens	proopiomelanocortin	230	49
1903	gi190183	Homo sapiens	opiomelanocortin	230	49
1904	gi 11037117 gb AAG27485.1 AF194537.1	Homo sapiens	NAG13	180	53
1905	gi5360984	Homo sapiens	dJ228H13.1 (similar to Ribosomal protein L21e)	152	72
1905	AAB44126	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1571.	150	83

201

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1905	gi550015	Homo sapiens	ribosomal protein L21	150	83
1906	gi2654610	Pseudomonas aeruginosa	arginine/ornithine succinyltransferase AI subunit	79	25
1906	gi17226812	Botryotinia fuckeliana	histidine kinase	72	33
1906	gi16904238	Botryotinia fuckeliana	two-component osmosensing histidine kinase BOS1p	72	33
1908	gi330359	Human herpesvirus 4	nuclear antigen precursor	91	37
1908	gi1632793	Human herpesvirus 4	EBNA3C (EBNA 4B) latent protein	91	37
1908	gi1184677	Candida albicans	hyphal wall protein 1	90	38
1909	gi13177635	Rattus norvegicus	phospholipase C beta-3	72	26
1909	gi1150880	Mus musculus	phospholipase C beta3	71	26
1909	gi17105044	Simian adenovirus 25	10.1 kDa	71	31
1910	gi9857054	Leishmania major	possible CG7055 protein	71	47
1910	gi1617560	Leishmania major	LCFACAS5; L5701.2	67	33
1910	gi9857054 emb CAC04011.1]	Leishmania major	possible CG7055 protein	71	47
1911	AAY87278	Homo sapiens	INCY- Human signal peptide containing protein HSPP-55 SEQ ID NO:55.	501	82
1911	AAB18912	Homo sapiens	GETH A novel polypeptide designated PRO1889.	501	82
1911	AAU27659	Homo sapiens	ZYMO Human protein AFP513481.	416	77
1912	gi2065210	Mus musculus	Pro-Pol-dUTPase polyprotein	434	80
1912	gi118676710 dbj BAB85007.1]	Homo sapiens	FLJ00254 protein	270	64
1913	gi5713196	Caenorhabditis elegans	liprin-alpha homolog SYD-2	479	38
1913	gi930343	Homo sapiens	LAR-interacting protein 1b	467	39
1913	gi930341	Homo sapiens	LAR-interacting protein 1a	467	39
1914	gi6651021	Mus musculus	semaphorin cytoplasmic domain-associated protein 3B	274	63
1914	gi6651019	Mus musculus	semaphorin cytoplasmic domain-associated protein 3A	274	63
1914	AAM25720	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1235.	266	61
1915	gi902214	Zea mays	RNA polymerase beta' subunit-2	72	24
1915	gi112482	Zea mays	RNA polymerase beta-2 subunit (AA 1-1527)	72	24
1915	gi11467184 ref NP_043017.1]	Zea mays	RNA polymerase beta' subunit-2	72	24
1916	gi1655432	Mus musculus	plexin 2	1135	58
1916	AAM93435	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3070.	1132	57
1916	gi961515	Xenopus laevis	plexin	1126	54



202

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1917	gi15559064	Mus musculus	SNAG1	86	38
1917	gi 20863586 ref XP_141581.1	Mus musculus	similar to dJ551D2.5 (novel protein)	88	30
1917	gi 18644890 ref NP_570614.1	Mus musculus	sorting nexin associated golgi protein 1	86	38
1918	gi19528383	Drosophila melanogaster	RE04404p	67	32
1919	AAM77461	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 37767.	189	79
1919	AAM64684	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 36789.	189	79
1919	gi 17477135 ref XP_063415.1	Homo sapiens	similar to embryonal stem cell specific gene 1	263	75
1920	gi2623757	Rattus norvegicus	neurabin	172	97
1920	gi2827450	Gallus gallus	KS5 protein	154	88
1920	gi13991829	Xenopus laevis	neurabin	145	83
1923	gi5532302	Heterocapsa triquetra	PSII CP47 apoprotein	75	29
1923	gi1881335	Bacillus subtilis	SIMILAR TO YQFU, YXKD, YITB OF B. SUBTILIS.	68	38
1923	gi 5532302 gb AAD44701.1	Heterocapsa triquetra	PSII CP47 apoprotein	75	29
1924	gi6855429	Leishmania major	possible mucin 1 precursor	77	33
1924	gi5832816	Caenorhabditis elegans	contains similarity to Pfam domain: PF01694 (Rhomboid family), Score=61.7, E-value=5.1e-15, N=1	74	34
1924	AAB51976	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 48 SEQ ID NO:108.	72	38
1925	AAB51635	Homo sapiens	ROSE/ Human secreted protein sequence encoded by gene 16 SEQ ID NO:75.	205	31
1925	AAB47128	Homo sapiens	INCY- CDIFF-6, Incyte ID No. 2009435CD1.	199	34
1925	ABB55766	Homo sapiens	FECH/ Human polypeptide SEQ ID NO 138.	197	38
1926	AAG89279	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 399.	330	44
1926	AAB70690	Homo sapiens	SREN- Human hDPP protein sequence SEQ ID NO:7.	319	44
1926	gi13182757	Homo sapiens	HTPAP	319	44
1927	gi13177290	Ectocarpus siliculosus virus	EsV-1-8	69	36
1928	gi18700171	Arabidopsis thaliana	AT5g20480/F7C8_70	86	39
1928	gi915207	Sus scrofa	gastric mucin	83	29

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1928	gi532113	Caenorhabditis elegans	homeotic region most like HMPB_DROME: homeotic proboscipedia protein	79	27
1929	ABB12295	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2665.	135	59
1929	AAG04080	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8161.	78	38
1929	gi9279807	Drosophila melanogaster	cortactin	77	27
1930	AAV81204_aal	Homo sapiens	GEHO Human CD7 cDNA.	872	73
1930	AAB36657	Homo sapiens	IMMV Human CD7 protein sequence SEQ ID NO:2.	872	73
1930	AAU02438	Homo sapiens	GEHO Human lymphocyte cell surface antigen CD7 polypeptide.	872	73
1931	gi2636248	Bacillus subtilis	similar to transaldolase (pentose phosphate)	73	29
1931	gi 21398633 ref NP_654618.1	Bacillus anthracis A2012	Transaldolase, Transaldolase [Bacillus	74	29
1931	gi 16080764 ref NP_391592.1	Bacillus subtilis	similar to transaldolase (pentose phosphate)	73	29
1932	AAB43545	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:990.	73	46
1932	AAM40234	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3379.	71	26
1934	gi3129962	Gallus gallus	B locus Lectin like Natural Killer cell surface protein	82	30
1934	AAB93791	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13545.	77	38
1934	gi2541864	Drosophila melanogaster	DAD polypeptide	77	32
1935	gi 4959869 gb AAD34536.1	Murine leukemia virus	polymerase	335	52
1935	gi 6524624 gb AAF15098.1	Phascolarctos cinereus	pol protein	331	52
1935	gi 9630313 ref NP_056790.1	Gibbon ape leukemia virus	pol polyprotein	328	52
1936	gi6562332	Arabidopsis thaliana	diaminopimelate decarboxylase	86	30
1936	gi7573355	Arabidopsis thaliana	diaminopimelate decarboxylase-like protein	86	30
1936	gi15146250	Arabidopsis thaliana	AT5g11880/F14F18_50	86	30
1939	AAU07442	Homo sapiens	GETH Human Wnt1 Upregulated protein 2 (WUP2).	300	100
1939	AAU07441	Homo sapiens	GETH Human Wnt1 Upregulated protein 1 (WUP1).	300	100
1939	AAB56802	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1380.	300	100

204

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1940	gi5802814	Homo sapiens	Gag-Pro-Pol-Env protein	587	57
1940	gi4185939	Human endogenous retrovirus K	pol protein	586	57
1940	gi5802821	Homo sapiens	Gag-Pro-Pol protein	586	57
1941	AAU83088	Homo sapiens	ZYMO Novel secreted protein Z2812G3P.	586	100
1941	AAB20275	Homo sapiens	SCHE Human interleukin DNAX 80.	535	76
1941	AAB20277	Homo sapiens	SCHE Human interleukin DNAX 80 variant.	529	76
1942	AAM06866	Homo sapiens	HYSE- Human foetal protein, SEQ ID NO: 1074.	994	100
1942	gi17426446	Homo sapiens	bA351K23.5 (novel protein)	933	54
1942	gi15099951	Mus musculus	diacylglycerol acyltransferase 2	915	55
1943	AAM06596	Homo sapiens	HYSE- Human foetal protein, SEQ ID NO: 327.	406	98
1943	gi15640499 ref NP_230126.1	Vibrio cholerae	S-adenosylmethionine synthase	67	51
1945	AAG75561	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6325.	327	100
1945	gi16416764	Homo sapiens	FKSG16	327	100
1945	gi13905212	Mus musculus	RIKEN cDNA 1200006F02 gene	261	79
1946	gi288174	Mus musculus	Oct2b	97	85
1946	gi53490	Mus musculus	Oct2.5 transcription factor	97	85
1946	gi9937478	Drosophila melanogaster	thyroid hormone receptor-associated protein TRAP170	72	39
1947	AAM66980	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27286.	170	69
1947	AAM54574	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26679.	170	69
1947	AAM75189	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 35495.	159	86
1948	AAV10874	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	100	100
1949	AAA27155_aal	Homo sapiens	GENE- Human P2 DNA.	100	100
1949	AAV94475	Homo sapiens	GENE- Predicted translation product of human P2 splice isoform, P2-B.	100	100
1949	AAV94474	Homo sapiens	GENE- Human P2 protein.	100	100
1950	gi9502082	Homo sapiens	tubby super-family protein	80	40
1950	gi9502080	Mus musculus	tubby super-family protein	77	41
1950	gi8118432	Oryza sativa	beta-expansin	73	35
1951	gi4808994	walleye epidermal hyperplasia virus type 1	envelope polyprotein	69	46
1951	gi15642893 ref NP_227934.1	Thermotoga maritima	ribonucleotide reductase, B12-dependent	66	46
1952	AAB80264	Homo sapiens	GETH Human PRO332 protein.	577	61

205

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1952	AAB33425	Homo sapiens	GETH Human PRO332 protein UNQ293 SEQ ID NO:57.	577	61
1952	AAY13396	Homo sapiens	GETH Amino acid sequence of protein PRO332.	577	61
1953	gi16648392	Drosophila melanogaster	LD39243p	449	61
1953	AAG73684	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4448.	371	55
1953	AAY48312	Homo sapiens	META- Human prostate cancer-associated protein 9.	371	55
1954	AAU84348	Homo sapiens	BAAK/ Protein MMP2 differentially expressed in breast cancer tissue.	2068	94
1954	ABB90738	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 208.	2068	94
1954	AAB84607	Homo sapiens	PFIZ Amino acid sequence of matrix metalloproteinase gelatinase A.	2068	94
1955	gi16769680	Drosophila melanogaster	LD46678p	245	35
1955	AAM66797	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27103.	148	80
1955	AAM54396	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26501.	148	80
1957	AAB80242	Homo sapiens	GETH Human PRO236 protein.	648	97
1957	AAM93378	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2955.	648	97
1957	AAB12157	Homo sapiens	PROT- Hydrophobic domain protein from clone HP03165 isolated from KB cells.	648	97
1958	AAM41696	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6627.	234	47
1958	AAU17119	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 684.	229	46
1958	gi16741621	Homo sapiens	Similar to RAB37, member of RAS oncogene family	228	47
1959	gi18025526	cercopithecine herpesvirus 15	LF3	140	30
1959	gi3153821	Mus musculus	plenty-of-prolines-101; POP101; SH3-philo-protein	137	25
1959	gi39255	Actinomyces viscosus	sialidase	129	28
1960	ABB12366	Homo sapiens	HYSE- Human bone marrow expressed protein SEQ ID NO: 120.	400	90
1960	AAO12936	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26828.	115	95
1960	AAM84898	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:12491.	113	82
1961	gi19110438	Homo sapiens	polycystin-1L1	190	94
1961	gi3115393	Rana pipiens	guanylate cyclase inhibitory protein	80	35
1961	gi3462887	Rattus norvegicus	alpha-fodrin	68	31
1962	AAU83130	Homo sapiens	ZYMO Novel secreted protein	1076	100

206

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			Z835892G6P.		
1962	gi1890354	Brassica napus	L-ascorbate peroxidase	80	33
1962	gi7529611	Leishmania major	hypothetical protein L787.06	79	31
1963	AAG78679	Homo sapiens	BODE- Human thrombotic protein 46.	467	86
1963	AAV87347	Homo sapiens	INCY- Human signal peptide containing protein HSPP-124 SEQ ID NO:124.	467	86
1963	AAB01431	Homo sapiens	MILL- Human TANGO 224 (form 2).	467	86
1964	gi3413504	Rattus norvegicus	Bassoon	81	26
1964	gi330452	human herpesvirus 5	DNA polymerase	79	28
1964	AAV69717_aa1	Homo sapiens	LUDW- Tumour rejection antigen precursor MAGE-C1 cDNA.	73	33
1965	gi 2323287 gb AAB6652.8.1	multiple sclerosis associated retrovirus	polyprotein	286	64
1965	gi 2351212 db BAA2206.4.1	Friend murine leukemia virus	gag-pol polyprotein (precursor protein)	179	47
1965	gi 9629516 ref NP_044738.1	Rauscher murine leukemia virus	Pol	179	47
1966	gi 2323287 gb AAB6652.8.1	multiple sclerosis associated retrovirus	polyprotein	476	65
1966	gi 2281588 gb AAB6416.0.1	synthetic construct	Pol	323	51
1966	gi 9626961 ref NP_057933.1	Murine leukemia virus	Pr180	323	51
1967	gi2065210	Mus musculus	Pro-Pol-dUTPase polyprotein	518	73
1967	AAM65715	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26021.	464	69
1967	AAM53338	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25443.	464	69
1968	AAG78149	Homo sapiens	BODE- Human polypeptide-cytochrome b5-13.	388	82
1968	gi3150438	Human endogenous retrovirus K	pol-env	345	55
1968	gi1469243	Human endogenous retrovirus K	pol/env	345	55
1969	gi21113108	Xanthomonas campestris pv. campestris str. ATCC 33913	TonB-dependent receptor	78	31

207

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1969	gi476274	Homo sapiens	R kappa B	77	23
1969	gi4206769	Acanthamoeba castellanii	myosin I heavy chain kinase	76	27
1970	gi 13310191 gb AAK18189.1 AF331500_1	multiple sclerosis associated retrovirus element	recombinant envelope protein	244	77
1970	gi 8272468 gb AAF74215.1 AF156963_1	Homo sapiens	envelope protein	219	81
1970	gi 21103962 gb AAM33141.1	Homo sapiens	enverin-2	219	77
1971	AAU83621	Homo sapiens	GETH Human PRO protein, Seq ID No 60.	320	100
1971	AAO05826	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 19718.	295	93
1971	AAM39560	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2705.	194	56
1972	gi6456112	Mus musculus	F-box protein FBX15	128	44
1972	gi21428946	Drosophila melanogaster	GH22104p	74	31
1972	gi 6456112 gb AAF09139.1	Mus musculus	F-box protein FBX15	128	44
1973	gi148270	Escherichia coli	lambda-integrase	550	94
1973	gi1790244	Escherichia coli K12	site-specific recombinase, acts on cer sequence of ColE1, effects chromosome segregation at cell division	550	94
1973	gi13364217	Escherichia coli O157:H7	site-specific recombinase XerC	544	92
1974	gi1805552	Escherichia coli	FORMATE HYDROGENLYASE TRANSCRIPTIONAL ACTIVATOR.	887	88
1974	gi1616960	Escherichia coli	HyfR	887	88
1974	gi7920396	Salmonella typhimurium	formate hydrogenlyase activator protein	522	54
1975	gi409795	Escherichia coli	No definition line found	1175	99
1975	gi15074592	Sinorhizobium meliloti	HYPOTHETICAL TRANSMEMBRANE PROTEIN	378	33
1975	gi17740718	Agrobacterium tumefaciens str. C58 (U. Washington)	Na+/Pi-cotransporter	372	34
1976	AAB82047	Homo sapiens	IGAK- Human mast cell surface antigen.	163	23
1976	gi12654783	Homo sapiens	Similar to loss of heterozygosity, 11, chromosomal region 2, gene A	163	23
1976	AAZ45690_aal	Homo sapiens	REGC cDNA sequence encoding the human minor vault protein p193.	108	25
1977	ABB56523	Homo sapiens	MERI Human NMDA receptor subunit SEQ ID NO 44.	73	28

208

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1977	AAW87504	Homo sapiens	SIBI- Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA24.	73	28
1978	AAG00471	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4552.	285	93
1978	gi298489	Papio hamadryas	SP-10	133	34
1978	gi452582	Vulpes vulpes	fox sperm acrosomal protein FSA-Acr.1	132	34
1979	AAB87128	Homo sapiens	MILL- Human secreted protein MANGO 349, SEQ ID NO:130.	490	86
1979	AAB87179	Homo sapiens	MILL- Human secreted protein MANGO 349 I21K variant, SEQ ID NO:227.	488	85
1979	AAB87181	Homo sapiens	MILL- Human secreted protein MANGO 349 E41D variant, SEQ ID NO:231.	487	85
1982	AAM75035	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 35341.	109	67
1982	AAM62231	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 34336.	109	67
1982	gi11967423	Mus musculus	vomeroneasal receptor V1RC5	105	76
1983	AAG89276	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 396.	224	46
1983	AAB56565	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1143.	99	40
1983	AAY44987	Homo sapiens	INCY- Human epidermal protein-4.	78	28
1984	AAB95089	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17025.	498	97
1984	AAM06608	Homo sapiens	HYSE- Human foetal protein, SEQ ID NO: 339.	495	96
1984	gi497890	unidentified nitrogen-fixing bacteria	alpha subunit of dinitrogenase reductase (Fe protein)	73	24
1985	gi117455728 refXP_063594.1	Homo sapiens	similar to Zinc-finger protein ubi-d4 (Requiem) (Apoptosis response zinc finger protein)	71	37
1986	gi21428886	Drosophila melanogaster	GH12469p	69	34
1987	gi7767529	Bos taurus	cyclophilin I	364	75
1987	gi8699209	Canis familiaris	cyclophilin A	361	88
1987	gi11641132	Sus scrofa	cyclophilin	361	88
1988	gi15073168	Sinorhizobium meliloti	PROBABLE TRANSLATION INITIATION FACTOR IF-2 PROTEIN	81	37
1988	gi1181352	Paramecium bursaria Chlorella virus 1	Pro-rich protein; PIPG (8X)	78	25
1988	gi493242	Feline herpesvirus 1	Feline herpesvirus type 1 immediate early protein	77	20
1989	AAM65707	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26013.	134	66

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1989	AAM53330	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25435.	134	66
1989	gi 20475216 ref XP_114802.1	Homo sapiens	similar to synapsin I	228	59
1990	AAM71181	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 31487.	110	64
1990	AAM58674	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 30779.	110	64
1990	gi21323636	Corynebacterium glutamicum ATCC 13032	Sulfate permease and related transporters (MFS superfamily)	75	26
1991	gi1932813	Xenopus laevis	dsRNA adenosine deaminase	96	34
1991	AAE10203	Homo sapiens	HYSE- Human bone marrow derived contig protein, SEQ ID NO: 68.	83	25
1991	gi3242649	Rana catesbeiana	alpha 1 type I collagen	80	30
1992	gi1181423	Paramecium bursaria Chlorella virus 1	PBCV-1 chitinase	71	41
1992	gi 21300897 gb EAA13042.1	Anopheles gambiae str. PEST	agCP14405	72	37
1992	gi 9631828 ref NP_048613.1	Paramecium bursaria Chlorella virus 1	PBCV-1 chitinase	71	41
1994	gi8248755	Plasmodium falciparum 3D7	protein phosphatase	72	25
1994	gi4104348	Campylobacter rectus	S-layer-RTX protein	70	38
1994	gi 8248755 emb CAB62878.2	Plasmodium falciparum 3D7	protein phosphatase	72	25
1995	gi21324402	Corynebacterium glutamicum ATCC 13032	Uncharacterized ATPase related to the helicase subunit of the Holliday junction resolvase	73	38
1995	gi 19552845 ref NP_600847.1	Corynebacterium glutamicum	COG2256:Uncharacterized ATPase related to the helicase subunit of the Holliday junction resolvase	73	38
1995	gi 17533213 ref NP_495777.1	Caenorhabditis elegans	F14E5.5.p	73	30
1996	gi1871223	Rickettsia typhi	crystalline surface layer protein	92	30
1996	gi6969926	Rickettsia aeschlimannii	OmpB	79	25
1996	gi14670347	Rickettsia felis	OmpB	78	25
1997	gi 20548733 ref XP_055641.2	Homo sapiens	similar to gag protein	256	58
1997	gi 9739120 gb AAF97916.1	Bovine leukemia virus	gag	186	34



210

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
1997	gi 9626226 ref NP_056897.1	Bovine leukemia virus	Pr44	185	34
1998	AAM79834	Homo sapiens	HYSE- Human protein SEQ ID NO 3480.	279	71
1998	AAM78850	Homo sapiens	HYSE- Human protein SEQ ID NO 1512.	279	71
1998	AAM79204	Homo sapiens	HYSE- Human protein SEQ ID NO 1866.	272	71
1999	AAM73176	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 33482.	168	48
1999	AAM60521	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 32626.	168	48
1999	gi 13929148 ref NP_113997.1	Rattus norvegicus	cyclic nucleotide-gated channel beta subunit 1	163	47
2000	gi1869859	human herpesvirus 2	very large tegument protein	73	30
2000	gi7380253	Neisseria meningitidis Z2491	2-keto-4-hydroxyglutarate aldolase	70	37
2000	gi7226633	Neisseria meningitidis MC58	4-hydroxy-2-oxoglutarate aldolase/2-dehydro-3-deoxyphosphogluconate aldolase	70	37
2001	gi17016969	Mus musculus	NUANCE	138	36
2001	gi6273778	Homo sapiens	trabeculin-alpha	137	33
2001	gi1675222	Mus musculus	ACF7 neural isoform 1	136	42
2002	AAM39256	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2401.	81	29
2002	gi840789	Homo sapiens	binding regulatory factor	81	29
2002	gi17028337	Homo sapiens	regulatory factor X, 5 (influences HLA class II expression)	81	29
2003	gi2252814	Mus musculus	FOG	172	64
2003	AAR58815	Homo sapiens	USSH Human c-myc far upstream element (FUSE) binding protein (FBP) variant from HL60 clone 3-1.	103	42
2003	gi3598974	Rattus norvegicus	protein tyrosine phosphatase TD14	103	26
2004	gi11994696	Arabidopsis thaliana	contains similarity to DNA repair protein-gene id:K7M2.11	77	28
2004	gi7209527	Mus musculus	testis-specific gene	73	24
2004	gi 17451912 ref XP_071083.1	Homo sapiens	similar to DNA-binding protein B	234	97
2005	AAE12023	Homo sapiens	INCY- Human G-protein coupled receptor, GCRC-2.	173	100
2005	AAG65832	Homo sapiens	FARB Human G protein-coupled receptor (GPCR).	173	100
2005	AAG68126	Homo sapiens	FARB Human 7TM-GPCR protein sequence SEQ ID NO:6.	105	78
2006	gi20068811	Homo sapiens	Rab-coupling protein	130	43
2006	gi15822596	Homo sapiens	nRip11	104	45

211

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
2006	gi13377897	Homo sapiens	Rab11 interacting protein Rip11a	83	40
2007	gi 17539708 ref NP_501489.1	Caenorhabditis elegans	F08B4.5.p	78	42
2008	AAE10350	Homo sapiens	PFIZ Human ADAMTS-J1.4 variant protein.	504	97
2008	AAE10349	Homo sapiens	PFIZ Human ADAMTS-J1.3 variant protein.	504	97
2008	AAE10347	Homo sapiens	PFIZ Human ADAMTS-J1.1 variant protein.	504	97
2009	AAV31720_aa1	Homo sapiens	MOUN Nucleotide sequence of the PUR-alpha gene.	87	29
2009	AAT99264_aa1	Homo sapiens	MOUN Human PUR-alpha gene.	87	29
2009	AAQ44800_aa1	Homo sapiens	MOUN Encodes single-stranded DNA binding (PUR) protein.	87	29
2010	gi170444	Lycopersicon esculentum	extensin (class II)	123	27
2010	gi4662641	Arabidopsis thaliana	expressed protein	116	30
2010	gi188864	Homo sapiens	mucin	115	28
2011	AAAY93650	Homo sapiens	HUMA- Amino acid sequence of a human prostacyclin-stimulating factor-2.	1677	100
2011	AAS15723_aa1	Homo sapiens	CURA- DNA encoding insulin-like growth factor family related protein, NOV3.	1673	99
2011	AAE17599	Homo sapiens	INCY- Human extracellular messenger (XMES)-1 protein.	1673	99
2012	gi10440434	Homo sapiens	FLJ00052 protein	336	69
2012	gi20502870	Mus musculus	SDS3	333	68
2012	gi21430678	Drosophila melanogaster	RE74901p	170	36
2013	AAH77293_aa1	Homo sapiens	MILL- Human ion channel protein IC32391 cDNA coding region.	214	93
2013	AAE13278	Homo sapiens	INCY- Human transporters and ion channels (TRICH)-5.	214	93
2013	AAG77969	Homo sapiens	MILL- Human ion channel protein IC32391.	214	93
2014	gi4894768	Xenopus laevis	ephrin-B2 precursor	78	30
2015	AAU77498	Homo sapiens	INCY- Human lipid metabolism enzyme, LMM-6.	1291	100
2015	ABB08205	Homo sapiens	INCY- Human lipid metabolism enzyme-5 (LME-5).	1122	100
2015	ABB07493	Homo sapiens	INCY- Human lipid metabolism molecule (LMM) polypeptide (ID: 2965233CD1).	864	75
2016	gi 14769015 ref XP_041569.1	Homo sapiens	fibrillin3	68	36
2017	gi2313786	Helicobacter pylori 26695	chorismate synthase (aroC)	78	33
2017	gi4155160	Helicobacter pylori J99	CHORISMATE SYNTHASE	72	32

212

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
2017	gi 15645287 refNP_207457.1	Helicobacter pylori 26695	chorismate synthase (aroC)	78	33
2018	gi15485622	Homo sapiens	Q9H4T4 like	1068	100
2018	ABB14744	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 3401.	694	98
2018	AAB95100	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17064.	101	24
2019	gi8050556	Gorilla gorilla	carboxyl-ester lipase	223	42
2019	AAU09894	Homo sapiens	MONS Bile Salt Stimulated Lipase (BSSL).	217	39
2019	ABB04676	Homo sapiens	MONS Human milk bile salt-stimulated lipase (BSSL) protein SEQ ID NO:2.	217	39
2020	gi2065210	Mus musculus	Pro-Pol-dUTPase polyprotein	515	74
2020	gi 385615 gb AAB26708.1	Mus sp.	fibulin gene homolog	300	75
2020	gi 13194728 gb AAK15526.1 AF329451.1	Gallus gallus	pol-like protein ENS-3	170	33
2021	AAM66980	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27286.	170	75
2021	AAM54574	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26679.	170	75
2021	AAM75189	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 35495.	159	86
2022	AAD29146_aa1	Homo sapiens	ZYMO Human Zcyto21 consensus cDNA.	649	83
2022	AAU83208	Homo sapiens	ZYMO Novel secreted protein Z908463G2P.	649	83
2022	AAE18311	Homo sapiens	ZYMO Human Zcyto21 consensus protein.	649	83
2024	gi14336750	Homo sapiens	Ce protein similar to Dm Cys3His finger protein	84	34
2024	AAB50363	Homo sapiens	UYSL- Human SRCAP.	83	34
2024	AAB95541	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18149.	83	34
2025	gi18676682	Homo sapiens	FLJ00240 protein	470	45
2025	gi14701866	Dictyostelium discoideum	carnil	221	29
2025	gi1881738	Acanthamoeba castellanii	myosin-I binding protein Acan125	219	29
2026	ABB12490	Homo sapiens	HYSE- Human bone marrow expressed protein SEQ ID NO: 329.	212	78
2027	AAU83147	Homo sapiens	ZYMO Novel secreted protein Z846363G2P.	1153	100
2027	gi 21287755 gb EAA00076.1	Anopheles gambiae str. PEST	ebiP4780	205	51

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
2027	gi 17552028 ref NP_498407.1	Caenorhabditis elegans	C05D11.8.p	91	38
2028	gi1510143	Homo sapiens	similar to C.elegans protein encoded in cosmid T20D3 (Z68220).	323	57
2028	gi3879942	Caenorhabditis elegans	T20D3.11	124	27
2028	gi5869818	Globodera pallida	NADH-ubiquinone oxidoreductase subunit 6	82	27
2029	AAE13288	Homo sapiens	INCY- Human transporters and ion channels (TRICH)-15.	75	31
2029	gi3252893	Thermotoga neapolitana	ABC transporter	74	37
2029	gi 18403965 ref NP_565826.1	Arabidopsis thaliana	expressed protein	70	29
2030	AAB97908	Homo sapiens	SHAN- Human GTP-binding protein 17 SEQ ID NO:2.	79	27
2030	AAM42129	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 7060.	79	27
2030	gi9971156	Mus musculus	GTP-binding like protein 2	79	27
2031	gi 20864803 ref XP_130800.1	Mus musculus	RIKEN cDNA 4930503K02	89	25
2031	gi 21262152 emb CAD32690.1	Oryza sativa	SMC4 protein	77	28
2031	gi 1507705 gb AAB06568.1	Borrelia burgdorferi	outer surface protein	74	33
2032	AAG65898	Homo sapiens	SMIK Amino acid sequence of GSK gene Id 18525.	481	100
2032	AAU83670	Homo sapiens	GETH Human PRO protein, Seq ID No 158.	471	97
2032	ABB84896	Homo sapiens	GETH Human PRO1309 protein sequence SEQ ID NO:160.	471	97
2034	gi6723273	Baboon endogenous virus strain M7	gag-pol precursor polyprotein	687	43
2034	gi18448744	Moloney murine leukemia virus	Pr180 gag-pro-pol polyprotein	685	42
2034	gi2801471	Moloney murine leukemia virus	Pr180	682	42
2035	gi 17554696 ref NP_497670.1	Caenorhabditis elegans	R148.7.p	68	32
2035	gi 16127996 ref NP_414543.1	Escherichia coli K12	aspartokinase I, homoserine dehydrogenase I	68	43
2035	gi 19548975 gb AAL90885.1 AF487900.1	Escherichia coli	aspartokinase I-homoserine dehydrogenase I	68	43
2036	gi13424459	Caulobacter	methyl-accepting chemotaxis protein	72	32

214

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		crescentus CB15	Mcpl		
2036	gi 16877133 gb AAH16838.1 AAH16838	Homo sapiens	carboxypeptidase, vitellogenic-like	69	30
2037	AAB67055	Homo sapiens	INCY- Human immune response molecule (IMUN) protein SEQ ID NO: 9.	532	75
2037	AAO01862	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15754.	403	67
2037	gi 6753924 ref NP_034374.1	Mus musculus	Friend virus susceptibility 1	240	39
2039	AAB38447	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 20 clone HUFBY15.	80	27
2039	gi11527799	Mus musculus	GTP-binding protein like 1	73	30
2039	gi695237	Equine herpesvirus 2	tegument protein	73	33
2040	gi 20544038 ref XP_089612.4	Homo sapiens	similar to PER-HEXAMER REPEAT PROTEIN 5	68	41
2042	AAM77922	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 38228.	642	85
2042	AAM65219	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 37324.	642	85
2042	gi 6723273 dbj BAA89659.1	Baboon endogenous virus strain M7	gag-pol precursor polyprotein	139	26
2043	gi48507	Wolinella succinogenes	formate dehydrogenase	80	27
2043	gi12381857	Danio rerio	c-Maf	78	42
2043	gi 18594822 ref XP_092995.1	Homo sapiens	zinc finger protein 21 (KOX 14)	306	100
2044	gi3132272	Sus scrofa	WT1 homologue	99	47
2044	AAG78446	Homo sapiens	MASI Predicted WT1 Wilm's tumour polypeptide of humans.	96	45
2044	AAG62154	Homo sapiens	CORI- Human WT1/PSA fusion protein SEQ ID NO: 357.	96	45
2046	gi21483222	Drosophila melanogaster	AT16994p	86	33
2046	gi21111736	Xanthomonas campestris pv. campestris str. ATCC 33913	cell division protein	79	30
2046	gi12653493	Homo sapiens	Similar to brain acid-soluble protein 1	79	36
2047	ABB12490	Homo sapiens	HYSE- Human bone marrow expressed protein SEQ ID NO: 329.	200	83
2047	gi 20837783 ref XP_145921.1	Mus musculus	similar to 40S ribosomal protein S11	73	35

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
2047	gi 6002932 gb AAF00209.1 AF164960.5	Streptomyces fradiae	glycosyl transferase	71	35
2048	AAB59012	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 720.	103	32
2048	gi2429362	Santalum album	proline rich protein	99	31
2048	gi17945382	Drosophila melanogaster	RE17165p	98	25
2051	gi15625542	Hepatitis B virus	S antigen	71	31
2051	gi 4884886 gb AAD31857.1 AF134140.1	Hepatitis B virus	surface antigen	68	30
2052	AAB28764	Homo sapiens	HUMA- Sequence homologous to protein fragment encoded by gene 21.	693	78
2052	gi2065210	Mus musculus	Pro-Pol-dUTPase polyprotein	693	78
2052	AAB73606	Homo sapiens	SHAN- Human dUTP pyrophosphatase 26.	668	77
2053	gi9945983	Pseudomonas aeruginosa	transcriptional regulator PcaQ	83	34
2053	gi13874427	Homo sapiens	cerebral protein-5	76	35
2053	gi12803205	Homo sapiens	CAAX box 1	76	35
2054	gi21307831	Aplysia californica	CREB-binding protein	76	26
2054	gi16755887	Drosophila melanogaster	guanine nucleotide exchange factor	76	26
2054	gi 21307831 gb AAL54859.1	Aplysia californica	CREB-binding protein	76	26
2055	gi16588389	Homo sapiens	B lymphocyte activation-related protein BC-1514	437	71
2055	AAB92981	Homo sapiens	HELL- Human protein sequence SEQ ID NO:11698.	407	68
2055	AAM48325	Homo sapiens	SHAN- Human purine receptor 21.23.	398	74
2056	gi 2072969 gb AAC51274.1	Homo sapiens	p40	134	47
2056	gi 7959889 gb AAF71115.1 AF116721.95	Homo sapiens	PRO2221	123	43
2056	gi 2072974 gb AAC51277.1	Homo sapiens	p40	122	44
2057	gi19171178	Homo sapiens	metalloprotease disintegrin 16 with thrombospondin type I motif	518	98
2057	gi19171150	Homo sapiens	ADAMTS18 protein	168	35
2057	AAM39212	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2357.	128	76
2058	gi 4959869 gb AAD34536.1	Murine leukemia virus	polymerase	336	50

216

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
2058	gi 9630313 ref NP_056790.1	Gibbon ape leukemia virus	pol polyprotein	331	46
2058	gi 6723273 dbj BAA89659.1	Baboon endogenous virus strain M7	gag-pol precursor polyprotein	329	49
2059	gi 20546404 ref XP_116466.1	Homo sapiens	similar to nuclear receptor coactivator 4; RET-activating gene ELE1	179	91
2060	gi 6731237 gb AAAF27177.1 AF182317.1	Homo sapiens	myoferlin	112	79
2060	gi 798799 gb AAC37713.1	Mus musculus	immunoglobulin heavy chain	72	55
2060	gi 20819487 ref XP_145357.1	Mus musculus	similar to LYRIC	72	27
2061	gi 415738	Euglena gracilis	PSII D1-polypeptide	75	27
2061	gi 11491	Euglena gracilis	32 kd protein	75	27
2061	gi 11488	Euglena gracilis	32-Kda thylakoid membrane protein	75	27
2062	gi 21360549	Arabidopsis thaliana	AT3g01480/F4P13_3	79	29
2062	gi 3337366	Arabidopsis thaliana	nodulin-like protein	68	36
2063	gi 7959778	Homo sapiens	PRO1546	121	42
2063	AAG02639	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6720.	119	53
2063	AAG02753	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6834.	110	45
2064	gi 15077406	Antheraea yamamai	fibroin	109	30
2064	AAB82806	Homo sapiens	BOST- Human low density lipoprotein binding protein 2 (LBP-2).	92	24
2064	AAO01059	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 14951.	90	30
2065	gi 200964	Mus musculus	serine 2 ultra high sulfur protein	80	30
2065	gi 200962	Mus musculus	serine 1 ultra high sulfur protein	80	30
2065	AAM99918	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 34.	75	28
2066	gi 544724	Cavia	cholecystokinin A receptor; CCK-A receptor	69	29
2066	gi 2541920	Rattus norvegicus	cholecystokinin type-A receptor	69	29
2066	gi 2114152	Mus musculus	cholecystokinin type-A receptor	69	29
2067	gi 2828586	Pongo pygmaeus	BRCA1	73	22
2068	AAM40813	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5744.	75	29
2068	AAM39027	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2172.	75	29
2068	AAAY25768	Homo sapiens	HUMA- Human secreted protein encoded from gene 58.	75	29
2070	gi 1334150	Mus musculus	unidentified reading frame (first ATG	169	28

217

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			at pos. 210)		
2070	gi557822	Saccharomyces cerevisiae	ma15, sta1, len: 1367, CAI: 0.3, AMYH_YEAST P08640 GLUCOAMYLASE S1 (EC 3.2.1.3)	133	20
2070	gi1304387	Saccharomyces cerevisiae var. diastaticus	glucoamylase	133	20
2071	gi17983056	Brucella melitensis	BETA-HEXOSAMINIDASE A	88	29
2071	gi1573917	Haemophilus influenzae Rd	multidrug resistance protein A (emrA)	81	33
2071	gi17982813	Brucella melitensis	NITROGEN REGULATION PROTEIN NTRB	80	26
2073	gi 17532255 ref NP_496431.1	Caenorhabditis elegans	ankyrin and proline rich domains	67	29
2074	gi19919730	Homo sapiens	BTEB5	704	97
2074	gi13195441	Homo sapiens	BTE-binding protein 4	478	64
2074	gi14549656	Mus musculus	dopamine receptor regulating factor	452	76
2076	AAE17482	Homo sapiens	ZYMO Human leucine-rich repeat-7 (ZLRR7) protein.	1326	100
2076	AAU83190	Homo sapiens	ZYMO Novel secreted protein Z887300G2P.	1326	100
2076	ABB11242	Homo sapiens	HYSE- Human SLIT-2 homologue, SEQ ID NO:1612.	568	99
2077	gi18893729	Pyrococcus furiosus DSM 3638	protease iv	74	34
2077	AAB94745	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15792.	71	34
2077	gi16413096	Listeria innocua	lin0656	68	35
2078	gi60675	Beet ringspot virus	polyprotein	75	37
2078	gi 14743288 ref XP_047191.1	Homo sapiens	similar to Alu subfamily J sequence contamination warning entry	92	58
2078	gi 20260801 ref NP_620113.1	Beet ringspot virus	polyprotein	75	37
2079	gi3834629	Mus musculus	diaphanous-related formin; p134 mDia2	208	67
2079	AAG74400	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5164.	71	36
2079	gi3171906	Homo sapiens	DIA-156 protein	71	36
2080	gi17298315	Homo sapiens	candidate tumor suppressor protein	125	100
2080	gi7861733	Homo sapiens	low density lipoprotein receptor related protein-deleted in tumor	125	100
2080	gi8926243	Mus musculus	low density lipoprotein receptor related protein LRP1B/LRP-DIT	90	63
2081	gi4574224	Fundulus heteroclitus	multidrug resistance transporter homolog	343	55
2081	gi16304396	Pseudopleuronectes americanus	multidrug resistance transporter-like protein	340	52
2081	gi3355757	Gallus gallus	ABC transporter protein	328	53



218

Table 2

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
2082	gi7532975	bacteriophage phi-8	P10	67	27

Table 3

SEQ ID NO:	Database entry ID	Description	*Results
1059	BL00349	CTF/NP-I proteins.	BL00349H 15.70 9.710e-09 8-45
1061	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 6.143e-10 29-61 DM00215 19.43 8.322e-09 40-72
1062	DM01354	kw TRANSCRIPTASE REVERSE II ORF2.	DM01354U 12.24 6.092e-12 80-99
1063	PR00944	COPPER ION BINDING PROTEIN SIGNATURE	PR00944E 9.18 7.132e-09 33-46
1076	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 9.217e-09 23-35
1089	PR00308	TYPE I ANTIFREEZE PROTEIN SIGNATURE	PR00308C 3.83 8.754e-10 16-25
1089	PR00456	RIBOSOMAL PROTEIN P2 SIGNATURE	PR00456E 3.06 9.658e-09 16-30
1089	PR00341	PRION PROTEIN SIGNATURE	PR00341E 3.32 9.898e-09 24-43
1099	PR00886	HIGH MOBILITY GROUP (HMG1/HMG2) PROTEIN SIGNATURE	PR00886C 11.84 1.141e-12 28-46
1107	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 3.077e-09 51-65
1118	BL00472	Small cytokines (intercrine/chemokine) C-C subfamily signatur.	BL00472A 7.45 5.655e-09 1-12
1118	PR00655	AUXIN BINDING PROTEIN SIGNATURE	PR00655E 8.06 9.000e-09 88-103
1119	BL00970	Nuclear transition protein 2 proteins.	BL00970C 14.80 8.183e-12 99-136
1119	BL00826	MARCKS family proteins.	BL00826B 12.51 4.279e-09 92-143
1119	BL00348	p53 tumor antigen proteins.	BL00348F 23.19 5.881e-10 93-135 BL00348F 23.19 6.857e-09 91-133
1119	PD01457	RIBOSOMAL PROTEIN 40S ZINC-FINGER METAL.	PD01457A 16.51 8.216e-09 73-117
1119	BL00752	XPA protein.	BL00752B 19.17 7.866e-09 100-143 BL00752B 19.17 8.979e-09 63-106
1119	DM01269	303 kw ACTIVATING RAN GTPASE ISOZYME.	DM01269A 23.35 9.446e-09 109-136
1124	DM01813	EGG-LAYING HORMONE.	DM01813A 15.31 5.215e-09 15-42
1127	BL00452	Guanylate cyclases proteins.	BL00452A 17.52 1.170e-09 6-27
1131	BL00113	Adenylate kinase proteins.	BL00113B 20.49 9.897e-09 157-200
1162	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 7.000e-35 24-62
1163	BL00407	Connexins proteins.	BL00407B 14.23 9.775e-30 21-51 BL00407C 14.61 2.500e-24 52-79
1163	PR00206	CONNEXIN SIGNATURE	PR00206B 13.75 1.957e-24 33-55 PR00206A 11.35 6.559e-23 2-26 PR00206C 15.16 7.469e-20 58-78
1171	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 8.500e-28 35-73
1177	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803C 7.00 7.240e-09 46-55
1190	PR00774	GUANYLIN PRECURSOR SIGNATURE	PR00774A 6.49 8.579e-10 69-81
1195	PD02059	CORE POLYPROTEIN PROTEIN GAG CONTAINS: P.	PD02059C 21.58 8.031e-09 100-140
1197	BL00472	Small cytokines (intercrine/chemokine) C-C subfamily signatur.	BL00472A 7.45 8.000e-14 1-12
1213	PR00437	SMALL CXC CYTOKINE	PR00437C 14.85 1.310e-16 33-51

Table 3

SEQ ID NO:	Database entry ID	Description	*Results
		FAMILY SIGNATURE	
1213	BL00471	Small cytokines (intercrine/chemokine) C-x-C subfamily signat.	BL00471 23.92 7.960e-10 6-53
1216	PR00308	TYPE I ANTIFREEZE PROTEIN SIGNATURE	PR00308C 3.83 5.208e-09 183-192
1222	PF00852	Fucosyl transferase.	PF00852F 15.97 1.409e-15 195-231
1224	BL00299	Ubiquitin domain proteins.	BL00299 28.84 6.301e-11 47-98
1230	PR00540	MUSCARINIC M3 RECEPTOR SIGNATURE	PR00540A 10.24 7.174e-09 134-153
1240	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 7.480e-10 160-182 BL00290B 13.17 2.875e-09 226-243
1258	PR00792	PEPSIN (A1) ASPARTIC PROTEASE FAMILY SIGNATURE	PR00792A 11.54 5.500e-18 80-100
1258	BL00141	Eukaryotic and viral aspartyl proteases proteins.	BL00141A 12.10 4.789e-15 87-102 BL00141B 12.14 2.929e-10 228-239
1300	BL00616	Histidine acid phosphatases phosphohistidine proteins.	BL00616A 11.86 1.000e-09 136-143
1301	DM01417	6 kw INDUCING XPMC2 MUSHROOM SPAC22G7.04.	DM01417C 12.93 9.325e-12 361-372 DM01417D 11.08 9.820e-12 400-415
1302	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 6.067e-11 324-338
1311	BL00926	Lysyl oxidase copper-binding region proteins.	BL00926B 13.84 7.453e-09 84-121
1320	PR00830	ENDOPEPTIDASE LA (LON) SERINE PROTEASE (S16) SIGNATURE	PR00830A 8.41 3.712e-09 29-48
1325	BL00048	Protamine P1 proteins.	BL00048 6.39 4.671e-10 58-84 BL00048 6.39 4.908e-10 60-86 BL00048 6.39 2.913e-09 59-85 BL00048 6.39 5.950e-09 57-83
1345	PF00424	REV protein (anti-repression transactivator protein).	PF00424A 14.34 2.436e-09 184-215
1345	BL00048	Protamine P1 proteins.	BL00048 6.39 4.553e-10 178-204 BL00048 6.39 6.513e-09 179-205
1353	DM01354	kw TRANSCRIPTASE REVERSE II ORF2.	DM01354U 12.24 2.857e-15 82-101
1363	PF00850	Histone deacetylase family.	PF00850B 10.13 5.154e-14 95-109 PF00850C 14.55 9.063e-11 132-148
1389	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 6.423e-09 50-64
1389	PD00306	PROTEIN GLYCOPROTEIN PRECURSOR RE.	PD00306B 5.57 7.000e-09 59-69
1396	BL00427	Disintegrins proteins.	BL00427 13.93 7.698e-17 260-314
1396	PR00289	DISINTEGRIN SIGNATURE	PR00289A 13.62 5.667e-14 274-293
1416	BL00419	Photosystem I psaA and psaB proteins.	BL00419B 22.23 9.489e-09 18-51
1434	PF00075	RNase H.	PF00075I 16.21 7.375e-11 167-173
1440	BL00598	Chromo domain proteins.	BL00598 14.45 1.500e-15 112-133
1440	PR00504	CHROMODOMAIN SIGNATURE	PR00504B 9.12 5.200e-13 106-120 PR00504C 11.19 6.510e-09 121-133
1450	PF00622	Domain in SPLa and the RYanodine Receptor.	PF00622B 21.00 2.227e-09 93-114
1451	PD02935	FATTY ACID OXIDOREDUCTASE BIOSYNT.	PD02935C 16.62 4.375e-16 59-86
1467	BL00479	Phorbol esters / diacylglycerol	BL00479A 19.86 3.000e-11 130-152

Table 3

SEQ ID NO:	Database entry ID	Description	*Results
		binding domain proteins.	BL00479B 12.57 3.340e-10 156-171
1468	PF00992	Troponin.	PF00992A 16.67 5.563e-10 139-173
1468	BL00795	Involucrin proteins.	BL00795C 17.06 3.600e-09 193-237
1468	PR00042	FOS TRANSFORMING PROTEIN SIGNATURE	PR00042D 8.97 7.554e-09 141-162
1474	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 9.308e-12 62-92
1474	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 1.563e-09 62-80
1474	BL00239	Receptor tyrosine kinase class II proteins.	BL00239C 18.75 4.205e-09 49-71
1475	BL00456	Sodium:solute symporter family proteins.	BL00456C 24.55 4.886e-28 15-69
1480	BL00983	Ly-6 / u-PAR domain proteins.	BL00983C 12.69 1.346e-09 36-51
1482	BL00979	G-protein coupled receptors family 3 proteins.	BL00979A 19.66 9.633e-12 74-121
1502	PD02561	DETHIOBIOTIN SYNTHETASE SYNTHASE.	PD02561B 12.71 9.308e-09 176-182
1506	BL00297	Heat shock hsp70 proteins family proteins.	BL00297H 15.46 9.625e-23 302-355 BL00297D 11.95 6.063e-21 166-205 BL00297E 18.56 6.077e-21 226-269 BL00297C 9.51 9.667e-15 105-156
1506	PR00301	70 KD HEAT SHOCK PROTEIN SIGNATURE	PR00301I 12.76 3.208e-11 320-336
1513	PR00130	DNASE I SIGNATURE	PR00130E 14.66 5.046e-09 237-266
1515	DM01242	3 THREONINE--TRNA LIGASE.	DM01242A 20.32 5.286e-20 163-206
1517	BL00983	Ly-6 / u-PAR domain proteins.	BL00983B 8.19 5.935e-10 40-49
1520	BL00415	Synapsins proteins.	BL00415P 2.37 3.914e-10 138-173
1520	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 3.746e-09 124-138 PR00049D 0.00 1.000e-08 123-137
1530	PF00075	RNase H.	PF00075F 12.87 5.500e-10 127-137
1537	PR00463	E-CLASS P450 GROUP I SIGNATURE	PR00463F 17.63 5.219e-13 288-306 PR00463A 11.40 8.714e-12 52-71 PR00463B 17.50 5.041e-10 76-97
1537	PR00385	P450 SUPERFAMILY SIGNATURE	PR00385C 16.94 6.318e-09 289-300
1538	PR00709	AVIDIN SIGNATURE	PR00709A 4.60 5.585e-09 19-37
1553	DM01354	kw TRANSCRIPTASE REVERSE II ORF2.	DM01354Y 10.69 6.423e-16 113-152
1558	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 6.400e-25 70-108
1564	PF00589	Phage integrase family.	PF00589B 16.17 1.621e-11 158-171 PF00589C 14.62 9.609e-10 183-194
1566	BL00908	Mandelate racemase / muconate lactonizing enzyme family signa.	BL00908B 37.71 6.455e-13 191-245
1567	PR00702	ACRIFLAVIN RESISTANCE PROTEIN FAMILY SIGNATURE	PR00702A 14.92 2.421e-25 8-32 PR00702B 12.77 9.690e-18 36-54
1570	BL01047	Heavy-metal-associated domain proteins.	BL01047A 13.50 5.125e-17 75-97
1575	DM01354	kw TRANSCRIPTASE REVERSE II ORF2.	DM01354U 12.24 9.429e-15 80-99
1606	PF00642	Zinc finger C-x8-C-x5-C-x3-H type (and similar).	PF00642 11.59 2.575e-11 197-207
1610	DM01354	kw TRANSCRIPTASE REVERSE II ORF2.	DM01354I 15.55 7.702e-34 348-388 DM01354G 11.57 3.625e-32 277-307 DM01354H 18.00 2.528e-23 308-347

Table 3

SEQ ID NO:	Database entry ID	Description	*Results
			DM01354F 14.56 4.088e-11 241-276
1616	PD02929	ADHESION GLYCOPROTEIN PRECURSOR I.	PD02929A 28.27 2.263e-25 32-85
1627	PR00121	SODIUM/POTASSIUM-TRANSPORTING ATPASE SIGNATURE	PR00121A 6.71 1.000e-08 15-29
1630	PR00824	HEPATIC LIPASE SIGNATURE	PR00824A 7.81 7.214e-22 6-24
1640	BL00359	Ribosomal protein L11 proteins.	BL00359C 22.18 1.155e-11 93-126
1641	PR00080	ALCOHOL DEHYDROGENASE SUPERFAMILY SIGNATURE	PR00080A 9.32 8.839e-10 134-145
1641	PR00081	GLUCOSE/RIBITOL DEHYDROGENASE FAMILY SIGNATURE	PR00081A 10.53 2.000e-12 45-62 PR00081E 17.54 1.783e-10 238-255 PR00081B 10.38 2.227e-09 134-145
1641	BL00061	Short-chain dehydrogenases/reductases family proteins.	BL00061A 9.41 9.053e-10 134-144 BL00061B 25.79 6.860e-09 197-234
1666	BL01257	Ribosomal protein L10e proteins.	BL01257D 18.80 2.973e-15 59-98
1667	BL01241	Link domain proteins.	BL01241 35.81 8.579e-37 180-232 BL01241 35.81 7.835e-14 289-341
1667	BL00086	Cytochrome P450 cysteine heme-iron ligand proteins.	BL00086 20.87 3.377e-09 283-314
1668	PR00671	INHIBIN BETA B CHAIN SIGNATURE	PR00671A 8.36 8.088e-09 4-22
1672	BL00674	AAA-protein family proteins.	BL00674E 15.24 5.680e-15 31-50
1682	PF00075	RNase H.	PF00075A 14.44 4.400e-13 73-89 PF00075C 11.58 8.442e-09 152-163
1689	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 6.471e-27 268-306
1689	PR00788	NITROPHORIN SIGNATURE	PR00788A 9.79 6.108e-09 3-15
1692	BL00299	Ubiquitin domain proteins.	BL00299 28.84 4.759e-10 32-83
1697	PR00423	CELL DIVISION PROTEIN FTSZ SIGNATURE	PR00423E 7.36 4.038e-09 20-41
1706	BL00795	Involucrin proteins.	BL00795C 17.06 5.395e-10 185-229
1709	BL00514	Fibrinogen beta and gamma chains C-terminal domain proteins.	BL00514C 17.41 3.618e-25 68-104 BL00514H 14.95 6.745e-16 230-254 BL00514G 15.98 6.566e-14 198-227 BL00514E 14.28 8.286e-14 128-144 BL00514D 15.35 2.915e-12 109-121
1714	PF00878	Cation-independent mannose-6-phosphate receptor repeat proteins.	PF00878T 17.51 3.818e-09 41-67
1715	PF01140	Matrix protein (MA), p15.	PF01140D 15.54 4.872e-09 123-157
1715	PF00992	Troponin.	PF00992A 16.67 6.451e-10 109-143 PF00992A 16.67 3.724e-09 98-132 PF00992A 16.67 6.684e-09 96-130
1718	PD02474	SYNTHASE SMALL SUBUNIT ACETOLACT.	PD02474B 21.08 7.940e-10 92-130
1725	BL00412	Neuromodulin (GAP-43) proteins.	BL00412B 10.60 1.000e-10 46-82
1725	PR00215	NEUROMODULIN SIGNATURE	PR00215C 13.98 6.116e-10 54-74
1725	DM01688	2 POLY-IG RECEPTOR.	DM01688G 16.45 3.160e-09 119-150 DM01688I 14.97 6.885e-09 107-154
1725	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 8.564e-09 303-335
1727	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 7.750e-21 185-215
1727	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 7.176e-12 185-203

Table 3

SEQ ID NO:	Database entry ID	Description	*Results
1727	BL00239	Receptor tyrosine kinase class II proteins.	BL00239B 25.15 4.387e-09 119-166
1728	BL00415	Synapsins proteins.	BL00415Q 2.23 8.115e-09 52-87
1734	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270B 22.18 5.567e-18 75-111 PD01270C 19.54 1.167e-17 118-146 PD01270A 17.22 4.960e-14 21-60 PD01270D 24.66 4.284e-09 152-187
1736	PD02346	PHOTOSYSTEM II PROTEIN PRECURSOR PHOTOSYNTHESIS.	PD02346A 9.24 8.851e-09 6-17
1741	BL00415	Synapsins proteins.	BL00415Q 2.23 6.777e-09 317-352
1744	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479B 12.57 1.000e-08 33-48
1750	PR00763	COAGULIN SIGNATURE	PR00763B 8.39 6.457e-09 41-60
1754	PR00276	INSULIN A CHAIN SIGNATURE	PR00276A 11.84 7.840e-09 46-55
1755	PR00042	FOS TRANSFORMING PROTEIN SIGNATURE	PR00042D 8.97 2.565e-09 164-185
1755	PF00922	Vesiculovirus phosphoprotein.	PF00922A 19.17 5.759e-09 99-132
1778	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 9.836e-14 59-80 PR00245C 7.84 1.540e-13 237-252 PR00245B 10.38 2.125e-13 176-190
1778	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 1.474e-12 90-129
1778	PR00534	MELANOCORTIN RECEPTOR FAMILY SIGNATURE	PR00534A 11.49 4.729e-09 51-63
1778	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237A 11.48 3.613e-09 26-50 PR00237C 15.69 7.525e-09 104-126
1787	PR00007	COMPLEMENT C1Q DOMAIN SIGNATURE	PR00007B 14.16 5.114e-15 146-165 PR00007A 19.33 7.052e-10 119-145
1787	PR00524	CHOLECYSTOKININ TYPE A RECEPTOR SIGNATURE	PR00524F 5.36 4.351e-09 70-83
1787	DM00250	kw ANNEXIN ANTIGEN PROLINE TUMOR.	DM00250B 13.84 6.595e-09 82-105
1787	BL00415	Synapsins proteins.	BL00415N 4.29 7.372e-09 62-105
1787	BL01113	C1q domain proteins.	BL01113B 18.26 3.786e-23 125-160 BL01113A 17.99 7.968e-15 73-99 BL01113A 17.99 5.091e-14 70-96 BL01113A 17.99 5.295e-11 64-90 BL01113A 17.99 8.568e-11 79-105 BL01113A 17.99 8.977e-11 67-93 BL01113A 17.99 4.635e-09 82-108 BL01113A 17.99 6.192e-09 76-102 BL01113A 17.99 7.750e-09 61-87
1787	BL00420	Speract receptor repeat proteins domain proteins.	BL00420A 20.42 8.691e-11 73-101 BL00420A 20.42 9.673e-11 70-98 BL00420A 20.42 2.180e-10 55-83 BL00420A 20.42 8.062e-09 52-80
1789	DM01930	2 kw FINGER SMCX SMCY YDR096W.	DM01930E 15.41 2.964e-33 45-89
1795	DM01688	2 POLY-IG RECEPTOR.	DM01688I 14.97 7.480e-10 107-154 DM01688J 14.69 4.455e-09 60-96
1796	PF00075	RNase H.	PF00075J 15.78 4.115e-13 115-132
1802	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 4.130e-11 86-98
1802	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 1.600e-10 110-126 BL00028 16.07 6.100e-10 70-86
1802	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048B 6.02 9.438e-10 83-92

Table 3

SEQ ID NO:	Database entry ID	Description	*Results
1812	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 4.130e-09 157-169
1824	PF00628	PHD-finger.	PF00628 15.84 5.500e-13 78-92
1833	PF00075	RNase H.	PF00075B 12.56 4.732e-10 156-166
1833	PR00939	C2HC-TYPE ZINC-FINGER SIGNATURE	PR00939A 8.95 3.045e-09 137-146
1842	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 3.192e-09 244-258
1844	BL00972	Ubiquitin carboxyl-terminal hydrolases family 2 proteins.	BL00972D 22.55 3.348e-11 168-192
1857	PF00424	REV protein (anti-repression transactivator protein).	PF00424A 14.34 8.085e-09 71-102
1860	PR00221	CAULIMOVIRUS COAT PROTEIN SIGNATURE	PR00221H 12.82 2.410e-09 184-197
1864	BL01282	BIR repeat proteins.	BL01282B 30.49 1.136e-10 214-252
1866	BL00155	Cutinase, serine proteins.	BL00155D 26.87 5.337e-09 19-67
1895	PF00075	RNase H.	PF00075F 12.87 7.353e-10 93-103
1911	BL00983	Ly-6 / u-PAR domain proteins.	BL00983C 12.69 6.365e-09 101-116
1911	BL00272	Snake toxins proteins.	BL00272C 8.27 1.000e-08 105-116
1925	PR00308	TYPE I ANTIFREEZE PROTEIN SIGNATURE	PR00308A 5.90 6.795e-11 64-78 PR00308C 3.83 2.385e-10 67-76
1925	PR00456	RIBOSOMAL PROTEIN P2 SIGNATURE	PR00456E 3.06 9.438e-10 57-71
1925	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 6.654e-09 59-73
1930	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 5.263e-10 107-116
1935	PF00075	RNase H.	PF00075J 15.78 2.309e-12 81-98
1940	PF00075	RNase H.	PF00075F 12.87 3.864e-09 74-84
1952	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 3.250e-10 184-197 PR00019A 11.19 5.667e-09 187-200
1954	BL00546	Matrixins cysteine switch.	BL00546A 19.62 8.105e-30 77-106
1954	BL00023	Type II fibronectin collagen-binding domain proteins.	BL00023 24.31 4.682e-35 340-376 BL00023 24.31 2.969e-28 282-318 BL00023 24.31 9.526e-24 224-260
1954	PR00138	MATRIXIN SIGNATURE	PR00138B 15.82 5.500e-18 144-159 PR00138A 15.14 8.773e-16 97-110
1954	BL00024	Hemopexin domain proteins.	BL00024B 21.53 9.591e-33 118-151 BL00024A 11.49 2.800e-13 97-107 BL00024C 22.98 7.796e-11 164-212
1954	PR00013	FIBRONECTIN TYPE II REPEAT SIGNATURE	PR00013C 12.29 1.000e-20 372-387 PR00013C 12.29 3.571e-15 314-329 PR00013C 12.29 7.800e-14 256-271 PR00013A 12.26 5.500e-13 344-353 PR00013B 14.75 1.237e-11 355-367 PR00013B 14.75 4.000e-09 297-309 PR00013A 12.26 5.333e-09 286-295 PR00013A 12.26 7.833e-09 228-237
1957	BL01182	Glycosyl hydrolases family 35 proteins.	BL01182A 21.39 3.357e-34 77-119
1957	PR00742	GLYCOSYL HYDROLASE FAMILY 35 SIGNATURE	PR00742B 15.52 2.653e-14 78-96 PR00742A 13.75 6.914e-10 57-74
1958	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449A 13.20 8.200e-15 214-235
1964	PR00727	BACTERIAL LEADER PEPTIDASE 1 (S26) FAMILY	PR00727A 12.93 7.000e-09 9-25

Table 3

SEQ ID NO:	Database entry ID	Description	*Results
		SIGNATURE	
1965	PF00075	RNase H.	PF00075D 10.71 7.188e-09 71-81
1966	PF00075	RNase H.	PF00075C 11.58 9.786e-11 110-121 PF00075B 12.56 1.878e-10 78-88
1968	DM00892	3 RETROVIRAL PROTEINASE.	DM00892C 23.55 4.082e-11 314-347
1970	PF00075	RNase H.	PF00075J 15.78 8.571e-10 335-352
1973	PF00589	Phage integrase family.	PF00589B 16.17 1.450e-14 101-114
1974	BL00675	Sigma-54 interaction domain proteins ATP-binding region A proteins.	BL00675B 24.07 1.000e-24 118-172 BL00675C 13.51 6.400e-24 183-210 BL00675D 12.03 1.750e-09 245-254
1987	PR00153	CYCLOPHILIN PEPTIDYL-PROLYL CIS-TRANS ISOMERASE SIGNATURE	PR00153B 11.57 1.500e-17 52-64 PR00153A 12.98 4.255e-10 23-38
1987	BL00170	Cyclophilin-type peptidyl-prolyl cis-trans isomerase signatur.	BL00170B 20.97 6.250e-33 47-86 BL00170A 17.08 2.309e-09 17-43
1998	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 7.750e-37 27-65 PD01066 19.43 8.863e-11 68-106
1999	PF00992	Troponin.	PF00992A 16.67 3.487e-09 108-142
1999	BL00224	Clathrin light chain proteins.	BL00224B 16.94 7.055e-09 96-148
1999	BL00422	Granins proteins.	BL00422C 16.18 8.059e-09 117-144
2001	BL00019	Actinin-type actin-binding domain proteins.	BL00019B 13.34 7.158e-14 261-283
2001	DM01354	kw TRANSCRIPTASE REVERSE II ORF2.	DM01354U 12.24 3.500e-13 345-364
2008	PD01719	PRECURSOR GLYCOPROTEIN SIGNAL RE.	PD01719A 12.89 3.483e-16 63-90
2011	BL00282	Kazal serine protease inhibitors family proteins.	BL00282 16.88 6.577e-10 127-149
2011	BL00222	Insulin-like growth factor binding proteins.	BL00222B 11.09 6.940e-10 74-89
2011	BL00621	Tissue factor proteins.	BL00621A 8.69 6.473e-09 5-22
2012	PD02563	PROTEIN NONSTRUCTURAL C VP18.	PD02563C 13.51 9.634e-10 74-128
2013	PR00124	ATP SYNTHASE C SUBUNIT SIGNATURE	PR00124A 8.81 5.655e-09 58-77
2013	PR00783	MAJOR INTRINSIC PROTEIN FAMILY SIGNATURE	PR00783C 13.54 8.981e-09 48-67
2034	PF00075	RNase H.	PF00075F 12.87 6.523e-09 183-193
2037	BL00326	Tropomyosins proteins.	BL00326D 8.76 9.327e-09 115-155
2048	PR00671	INHIBIN BETA B CHAIN SIGNATURE	PR00671B 4.29 8.767e-10 138-157
2052	PD02455	ELEMENT TRANSPOSABLE INSERTION PROTEIN TRANSPOSITION DNA.	PD02455C 29.23 5.230e-09 225-276
2058	PF00075	RNase H.	PF00075J 15.78 9.000e-10 81-98
2074	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 4.000e-13 62-74
2074	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048B 6.02 4.462e-11 59-68 PR00048B 6.02 1.000e-10 89-98 PR00048A 10.52 9.609e-10 101-114
2074	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 9.100e-13 104-120 BL00028 16.07 1.000e-08 46-62
2076	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 1.900e-11 106-119



226

## Table 3

\* Results include in order: Accession No., subtype, e-value, and amino acid position of the signature in the corresponding polypeptide

Table 4

SEQ ID NO:	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
1050	FAA_hydrolase	Fumarylacetoacetate (FAA) hydrolase fam	0.64	-89.1	1	22-143
1066	rubredoxin	Rubredoxin	7.2	-11.1	1	4-37
1076	ank	Ankyrin repeat	0.01	22.5	1	25-57
1076	sodfe_C	Iron/manganese superoxide dismutases, C-term	3.9	-67.9	1	38-124
1076	DUF232	Putative transcriptional regulator	8.1	-29.1	1	134-254
1099	HMG_box	HMG (high mobility group) box	8	-22.4	1	17-61
1109	UPAR_LY6	u-PAR/Ly-6 domain	0.21	-6.2	1	34-112
1110	ldl_recept_a	Low-density lipoprotein receptor domain	8.8e-07	36.0	1	196-240
1110	CUB	CUB domain	0.38	-27.8	1	52-161
1118	rvt	Reverse transcriptase	0.95	-46.1	1	38-207
1125	adenylatekinase	Adenylate kinase	0.00037	-77.6	1	13-103
1162	KRAB	KRAB box	1.1e-23	92.1	1	22-62
1163	connexin	Connexin	3.1e-23	90.6	1	1-130
1171	KRAB	KRAB box	6.6e-22	86.2	1	33-73
1193	MHC_I	Class I Histocompatibility antigen, domains	2e-06	1.1	1	29-205
1209	DOMON	DOMON domain	1.9e-12	54.8	1	102-215
1213	IL8	Small cytokines (intecrine/chemokine), inter	0.59	-7.8	1	18-55
1218	cys_rich_FGFR	Cysteine rich repeat	4.4	-11.0	1	28-76
1222	Glyco_transf_10	Glycosyltransferase family 10	6.6e-06	-54.1	1	1-322
1240	ig	Immunoglobulin domain	1.6e-06	35.1	2	41-124;156-230
1258	asp	Eukaryotic aspartyl protease	8e-06	-110.8	1	19-241
1280	DOMON	DOMON domain	8.9	-16.6	1	35-117
1288	PDZ	PDZ domain (Also known as DHR or GLGF)	1.1	0.4	1	7-73
1301	Exonuclease	Exonuclease	3.4e-33	123.7	1	322-479
1311	Gemini_mov	Geminivirus putative movement protein	5.7	-40.5	1	15-79
1341	fn3	Fibronectin type III domain	6.6e-36	132.7	2	109-200;212-301
1345	Collagen	Collagen triple helix repeat (20 copies)	7.3	-65.8	1	185-243
1365	Amidase	Amidase	0.017	-178.9	1	68-276
1375	Galactosyl_T	Galactosyltransferase	7.1e-44	159.2	1	113-309
1375	Glyco_transf_25	Glycosyltransferase family 25	3	-77.1	1	146-293
1381	GRAM	GRAM domain	6.6e-14	59.6	1	65-116
1396	Pep_M12B_prop ep	Reprolysin family propeptide	1.4e-27	105.1	1	75-191
1396	disintegrin	Disintegrin	2.6e-10	47.7	1	243-318
1398	SK_channel	Calcium-activated SK potassium channel	1.8e-06	34.9	1	1-57
1413	ig	Immunoglobulin domain	5.4	9.1	1	29-88
1416	dUTPase	dUTPase	0.00044	9.6	1	111-237
1420	Folate_rec	Folate receptor family	1.7	-111.2	1	14-175
1434	lectin_c	Lectin C-type domain	1.5e-05	28.0	1	233-319
1440	chromo	'chromo' (CHRromatin Organization MOdifier)	4.6e-11	50.2	1	92-133
1449	PMSR	Peptide methionine sulfoxide reductase	0.0089	-65.8	1	4-79
1450	SPRY	SPRY domain	9e-26	99.0	1	109-240

Table 4

SEQ ID NO:	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
1451	MaoC_dehydratases	MaoC like domain	2.1e-15	64.6	1	31-152
1463	NTP_transf_2	Nucleotidyltransferase domain	2.6e-12	54.3	1	121-234
1467	DAG_PE-bind	Phorbol esters/diacylglycerol binding domain	8.7e-05	27.4	1	130-180
1467	DC1	DC1 domain	0.66	11.2	1	141-172
1470	jmjC	jmjC domain	0.46	-18.2	1	166-262
1474	pkinase	Protein kinase domain	0.0019	-85.7	1	2-187
1475	SSF	Sodium:solute symporter family	0.13	-177.1	1	1-311
1478	dUTPase	dUTPase	7.6	-37.5	1	2-98
1479	fn3	Fibronectin type III domain	1.1e-19	78.9	1	14-100
1485	rnaseH	RNase H	0.36	-28.0	1	59-175
1488	NTR	NTR/C345C module	0.044	-6.1	1	293-398
1506	HSP70	Hsp70 protein	1.6e-13	38.3	1	61-424
1517	UPAR_LY6	u-PAR/Ly-6 domain	0.33	-8.2	1	44-106
1530	rnaseH	RNase H	0.011	-11.7	1	64-155
1537	p450	Cytochrome P450	2.1	-176.6	1	31-316
1537	DNA_ligase_OB	NAD-dependent DNA ligase OB-fold domain	9.2	-42.9	1	200-256
1558	KRAB	KRAB box	1.8e-18	74.8	1	68-108
1564	Phage_integrase	Phage integrase family	1.2e-09	45.5	1	39-204
1566	MR_MLE	Mandelate racemase / muconate lactonizing enzyme	0.00079	-24.5	1	153-352
1570	HMA	Heavy-metal-associated domain	6.6e-13	56.3	1	71-131
1580	ig	Immunoglobulin domain	0.99	15.2	1	23-131
1601	WD40	WD domain, G-beta repeat	2e-08	41.5	3	39-75:83-118:126-162
1606	zf-CCCH	Zinc finger C-x8-C-x5-C-x3-H type	0.094	19.3	3	105-129:141-173:183-209
1612	zf-CCHC	Zinc knuckle	2.1e-05	31.4	2	167-184:202-219
1618	rnaseH	RNase H	6.3e-14	59.7	1	24-144
1618	Integrase_Zn	Integrase Zinc binding domain	3.8e-07	37.2	1	146-185
1618	DUF224	Domain of unknown function (DUF224)	9.3	-7.0	1	104-186
1641	adh_short	short chain dehydrogenase	4.6e-32	119.9	1	42-309
1667	Xlink	Extracellular link domain	2.9e-83	290.0	2	162-267:273-364
1667	ig	Immunoglobulin domain	0.0015	25.2	1	61-145
1682	rvt	Reverse transcriptase	3.1e-31	117.2	1	56-238
1683	Gag_p30	Gag P30 core shell protein	2.9e-33	124.0	1	8-197
1689	KRAB	KRAB box	4.9e-22	86.6	1	266-306
1692	ubiquitin	Ubiquitin family	0.00061	26.5	1	17-91
1709	fibrinogen_C	Fibrinogen beta and gamma chains, C-term	7.9e-85	295.2	1	37-255
1713	HOK_GEF	Hok/gef family	2.4	-7.8	1	7-54
1716	Gag_p30	Gag P30 core shell protein	0.0036	-49.7	1	64-229
1721	rnaseH	RNase H	0.011	-11.7	1	207-350
1722	dUTPase	dUTPase	0.37	-22.9	1	93-217

Table 4

SEQ ID NO:	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
1725	ig	Immunoglobulin domain	4.2e-13	57.0	2	80-141:259-320
1725	IQ	IQ calmodulin-binding motif	4.3e-05	30.4	1	49-69
1727	pkinase	Protein kinase domain	3e-21	84.0	1	71-267
1728	Fringe	Fringe-like	5.9	-112.6	1	165-370
1734	ig	Immunoglobulin domain	0.014	22.0	1	117-170
1737	PP2C	Protein phosphatase 2C	0.0067	-50.5	1	37-273
1738	SH3	SH3 domain	1.7e-05	31.7	1	102-159
1740	maseH	RNase H	0.0042	-7.3	1	126-270
1744	DAG_PE-bind	Phorbol esters/diacylglycerol binding dom	2.9	-11.1	1	26-55
1744	PHD	PHD-finger	3.3	-14.7	1	9-61
1760	GARS_N	Phosphoribosylglycinamide synthetase, N	8.2	-62.0	1	35-95
1760	Armadillo_seg	Armadillo/beta-catenin-like repeat	9.1	8.7	2	44-84:131-171
1778	7tm_1	7 transmembrane receptor (rhodopsin family)	1e-12	55.7	1	41-276
1778	YCF9	YCF9	3.1	-18.5	1	203-258
1787	Clq	Clq domain	1e-05	13.2	1	111-230
1787	Collagen	Collagen triple helix repeat (20 copies)	0.0043	-3.0	1	50-107
1789	junJC	junJC domain	0.00078	12.0	1	52-241
1795	ig	Immunoglobulin domain	0.0037	23.9	1	64-141
1796	rve	Integrase core domain	2.6e-28	107.5	1	20-174
1802	zf-C2H2	Zinc finger, C2H2 type	6e-15	63.1	2	68-90:108-130
1806	Filamin	Filamin/ABP280 repeat	0.00054	18.6	1	26-131
1812	ank	Ankyrin repeat	3.6e-23	90.4	3	159-191:205-237:244-276
1824	PHD	PHD-finger	1.1e-12	55.6	1	62-110
1826	PAP assoc	PAP/25A associated domain	1.5e-06	35.2	1	101-155
1827	ig	Immunoglobulin domain	1.6	13.4	1	29-102
1830	RhoGEF	RhoGEF domain	3.3e-06	24.0	1	110-280
1830	PH	PH domain	2.8	6.7	1	356-451
1833	zf-CCHC	Zinc knuckle	2.1e-06	34.7	1	137-154
1833	rvt	Reverse transcriptase	7.7e-06	25.9	1	84-277
1844	UCH-2	Ubiquitin carboxyl-terminal hydrolase family	0.15	-8.5	1	165-238
1846	Armadillo_seg	Armadillo/beta-catenin-like repeat	0.28	17.7	2	50-91:92-132
1860	zf-CCHC	Zinc knuckle	3.2e-05	30.8	1	179-196
1864	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.0022	23.3	1	218-256
1887	ig	Immunoglobulin domain	4e-08	40.4	1	35-112
1889	LRR	Leucine Rich Repeat	0.051	20.1	1	62-85
1895	maseH	RNase H	3.4e-06	25.8	1	47-177
1899	Brevenin	Brevenin/esculentin/gaegurin/rugosin family	7.5	-2.9	1	1-51
1911	UPAR_LY6	u-PAR/Ly-6 domain	1.3e-06	35.4	1	44-117

Table 4

SEQ ID NO:	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
1911	toxin	Snake toxin	3	-19.5	1	66-117
1911	Activin_recp	Activin types I and II receptor domain	9.5	-14.0	1	30-118
1912	rvp	Retroviral aspartyl protease	7	-26.3	1	42-142
1913	SAM	SAM domain (Sterile alpha motif)	3.9e-13	57.1	2	105-170:183-247
1916	Sema	Sema domain	1.4e-14	54.6	1	51-434
1926	PAP2	PAP2 superfamily	2.9e-07	37.6	1	48-142
1930	ig	Immunoglobulin domain	2.7e-07	37.6	1	41-116
1935	rve	Integrase core domain	2.5e-13	57.7	1	1-138
1940	rnaseH	RNase H	1.1e-26	102.0	1	24-153
1940	Integrase_Zn	Integrase Zinc binding domain	4.7e-12	53.5	1	155-194
1952	LRRNT	Leucine rich repeat N-terminal domain	0.0027	24.4	1	67-95
1953	UQ_con	Ubiquitin-conjugating enzyme	2.8e-08	40.9	1	78-219
1954	Peptidase_M10	Matrixin	6.7e-86	298.8	1	53-212
1954	fn2	Fibronectin type II domain	1e-79	278.2	3	231-272:289-330:347-388
1958	ras	Ras family	1.9	-132.0	1	215-284
1963	tsp_1	Thrombospondin type 1 domain	0.083	8.0	1	20-63
1966	rvt	Reverse transcriptase	1.5e-05	21.9	1	2-196
1968	G-patch	G-patch domain	0.3	6.0	1	307-352
1968	rvp	Retroviral aspartyl protease	1.4	-19.9	1	274-385
1970	rve	Integrase core domain	0.78	-16.8	1	265-395
1973	Phage_integrase	Phage integrase family	5.7e-08	39.9	1	1-153
1974	Sigma54_activat	Sigma-54 interaction domain	3.1e-37	137.2	1	63-253
1975	Na_Pi_cotrans	Na <sup>+</sup> /Pi-cotransporter	0.0085	-99.2	1	1-146
1975	signal	His Kinase A (phosphoacceptor) domain	7	-7.7	1	85-147
1978	UPAR_LY6	u-PAR/Ly-6 domain	1.8	-16.0	1	21-96
1978	Zn_clus	Fungal Zn(2)-Cys(6) binuclear cluster domain	5.1	-5.7	1	21-60
1987	pro_isomerase	Cyclophilin type peptidyl-prolyl cis-tr	1.2e-18	75.4	1	4-171
1997	zf-CCHC	Zinc knuckle	1.9e-05	31.5	2	181-198:204-220
1997	TFIID-31	Transcription initiation factor IID, 31kD su	7.9	-63.3	1	75-187
1997	Gag_p12	Gag polyprotein, inner coat protein p12	8.9	-9.5	1	155-229
1998	KRAB	KRAB box	2e-23	91.2	1	27-65
2001	CH	Calponin homology (CH) domain	0.019	10.8	1	230-330
2001	SAM	SAM domain (Sterile alpha motif)	0.9	6.5	1	248-311
2008	tsp_1	Thrombospondin type 1 domain	0.013	15.1	1	64-98
2011	ig	Immunoglobulin domain	1.7e-05	31.7	1	186-255
2011	kazal	Kazal-type serine protease inhibitor domain	0.00028	27.6	1	121-168
2011	IGFBP	Insulin-like growth factor binding protein	0.17	2.5	1	53-113
2011	zf-UBR1	Putative zinc finger in N-recogin	8.3	-24.0	1	54-112
2015	PH	PH domain	0.0002	28.1	1	174-281
2015	efhand	EF hand	0.00031	27.5	1	339-367
2018	RPEL	RPEL repeat	1.3	11.8	1	25-50
2034	rnaseH	RNase H	4e-27	103.6	1	122-267

Table 4

SEQ ID NO:	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
2038	granulin	Granulin	7.7	-17.8	1	62-91
2052	rve	Integrase core domain	2.6e-24	94.2	1	160-314
2057	Pep_M12B_prop ep	Reprolysin family propeptide	0.44	-29.3	1	179-263
2058	rve	Integrase core domain	8.7e-14	59.2	1	1-140
2074	zf-C2H2	Zinc finger, C2H2 type	5.5e-22	86.5	3	42-66:72-96:102-124
2074	zf-BED	BED zinc finger	0.94	1.8	1	91-129
2074	TP1	Nuclear transition protein 1	7.5	2.2	1	21-76
2076	LRR	Leucine Rich Repeat	3.2e-20	80.6	5	57-80:81-104:105-128:129-152:153-176
2076	LRRNT	Leucine rich repeat N-terminal domain	0.00013	28.8	1	27-55
2076	LRRCT	Leucine rich repeat C-terminal domain	0.047	18.0	1	186-234

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1050	1qgj	A	16	52	3.4e-06	-0.68	0.41		FUMARYLACETOACETATE HYDROLASE; CHAIN: A, B;	HYDROLASE BETADIKETONASE, FAA; MIXED BETA-SANDWICH ROLL, HYDROLASE
1050	1qgj	A	16	54	1.3e-11	-0.70	0.42		FUMARYLACETOACETATE HYDROLASE; CHAIN: A, B;	HYDROLASE BETADIKETONASE, FAA; MIXED BETA-SANDWICH ROLL, HYDROLASE
1061	1ciu		34	172	9.6e-11	0.02	-0.19		CYCLODEXTRIN GLYCOSYLTRANSFERASE; 1CIU 6 CHAIN: NULL; 1CIU 7	GLYCOSIDASE CGTASE; 1CIU 8 THERMOSTABLE 1CIU 14
1061	1cww	A	50	226	9.6e-13	0.11	-0.19		INVASIN; CHAIN: A;	STRUCTURAL PROTEIN INTEGRIN-BINDING PROTEIN, INV GENE
1061	1pam	A	36	225	9.6e-15	0.21	-0.20		CYCLODEXTRIN GLUCANOTRANSFERASE; CHAIN: A, B;	GLYCOSYLTRANSFERASE TRANSFERASE, GLYCOSYLTRANSFERASE, CALCIUM, SIGNAL
1061	2tbv	C	39	170	9.6e-12	0.44	-0.19		VIRUS TOMATO BUSHY STUNT VIRUS 2TBV 4	
1076	1av1	A	60	257	3.2e-07			61.14	APOLIPOPROTEIN A-I; CHAIN: A, B, C, D;	LIPID TRANSPORT APO A-I; LIPOPROTEIN, LIPID TRANSPORT, CHOLESTEROL METABOLISM, 2 ATHEROSCLEROSIS, HDL, LCAT-ACTIVATION

Table 5

SEQ ID NO.	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEOFOLD score	Compound	PDB annotation
1076	1awc	B	19	82	9.6e-09	-0.24	0.15		GA BINDING PROTEIN ALPHA; CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABPALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA-BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
1076	1bd8		19	66	4.8e-07	-0.33	0.18		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
1076	1blx	B	19	66	2.4e-07	-0.32	0.27		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA, COMPLEX (INHIBITOR PROTEIN/KINASE)
1076	1bu9	A	19	66	4.3e-07	0.05	0.22		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A;	HORMONE/GROWTH FACTOR P18-INK4C; CELL CYCLE INHIBITOR, P18INK4C, TUMOR, SUPPRESSOR, CYCLIN-2 DEPENDENT KINASE, HORMONE/GROWTH FACTOR
1076	1cun	A	46	259	1.4e-05			72.19	ALPHA SPECTRIN; CHAIN: A, B, C;	STRUCTURAL PROTEIN TWO REPEATS OF SPECTRIN, ALPHA HELICAL LINKER REGION, 2 2



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1076	1myo		19	66	9.6e-09	-0.44	0.22		MYOTROPHIN; CHAIN: NULL	TANDEM 3-HELIX COILED-COILS, STRUCTURAL PROTEIN
1076	1qge	A	24	284	0.00096			62.82	VESICULAR TRANSPORT PROTEIN SEC17; CHAIN: A;	ANK-REPEAT MYOTROPHIN, ACETYLATION, NMR, ANK-REPEAT
1076	1quu	A	38	289	3.4e-05			68.38	HUMAN SKELETAL MUSCLE ALPHA-ACTININ 2; CHAIN: A;	PROTEIN TRANSPORT HELIX-TURN-HELIX TPR-LIKE REPEAT, PROTEIN TRANSPORT
1089	1aga		76	146	0.0048	0.24	0.15		CYTOCHROME B5; CHAIN: NULL;	CONTRACTILE PROTEIN TRIPLE-HELIX COILED COIL, CONTRACTILE PROTEIN
1089	1aw3		76	146	0.0038	0.22	0.10		CYTOCHROME B5; CHAIN: NULL;	ELECTRON TRANSPORT CYTOCHROME B5, PROTEIN RECOGNITION, SOLUTION STRUCTURES, 2 SECONDARY STRUCTURES, ELECTRON TRANSPORT
1089	1awp	A	76	148	0.0096	0.47	0.58		CYTOCHROME B5; CHAIN: A, B;	ELECTRON TRANSPORT CYTOCHROME, ELECTRON TRANSPORT, HEME
1089	1eyo		76	146	0.0048	0.44	0.10		ELECTRON TRANSPORT	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1089	1d69	A	76	151	0.0048	0.11	0.16		CYTOCHROME B5 (OXIDIZED) ICYO 3 ELECTRON TRANSPORT ICYO 4A	ELECTRON TRANSPORT CYTOCHROME, HEME
1099	1hme		8	61	3.2e-23	-0.62	0.98		DNA-BINDING HIGH MOBILITY GROUP PROTEIN FRAGMENT-B (HMG1) (DNA-BINDING 1HME 3 HMG-BOX DOMAIN B OF RAT HMG1) (NMR, 1 STRUCTURE) 1HME 4	
1099	1hsm		11	61	3.2e-21	-0.83	1.00		DNA-BINDING HIGH MOBILITY GROUP PROTEIN 1 (HMG1) BOX 2, COMPLEXED WITH 1HSM 3 MERCAPTOETHANOL (NMR, MINIMIZED AVERAGE STRUCTURE) 1HSM 4	
1102	1ezg	A	146	224	3.2e-08	1.04	-0.08		THERMAL HYSTERESIS PROTEIN ISOFORM YL-1; CHAIN: A, B;	ANTIFREEZE PROTEIN INSECT ANTIFREEZE PROTEIN, THERMAL HYSTERESIS, TENEBRIO 2 MOLITOR, IODINATION, RIGHT-HANDED BETA-HELIX, TMAFP
1102	1ezg	A	158	236	6.4e-10	1.04	-0.08		THERMAL HYSTERESIS PROTEIN ISOFORM YL-1;	ANTIFREEZE PROTEIN INSECT ANTIFREEZE PROTEIN,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN: A, B;	THERMAL HYSTERESIS, TENEBRIO 2 MOLITOR, IODINATION, RIGHT-HANDED BETA-HELIX, TMAFP
1102	1ezg	A	180	238	1.1e-10	0.28	-0.19		THERMAL HYSTERESIS PROTEIN ISOFORM YL-1; CHAIN: A, B;	ANTIFREEZE PROTEIN INSECT ANTIFREEZE PROTEIN, THERMAL HYSTERESIS, TENEBRIO 2 MOLITOR, IODINATION, RIGHT-HANDED BETA-HELIX, TMAFP
1102	1ezg	A	193	270	1.6e-11	0.80	-0.13		THERMAL HYSTERESIS PROTEIN ISOFORM YL-1; CHAIN: A, B;	ANTIFREEZE PROTEIN INSECT ANTIFREEZE PROTEIN, THERMAL HYSTERESIS, TENEBRIO 2 MOLITOR, IODINATION, RIGHT-HANDED BETA-HELIX, TMAFP
1102	1ezg	A	263	344	1.6e-10	1.13	0.10		THERMAL HYSTERESIS PROTEIN ISOFORM YL-1; CHAIN: A, B;	ANTIFREEZE PROTEIN INSECT ANTIFREEZE PROTEIN, THERMAL HYSTERESIS, TENEBRIO 2 MOLITOR, IODINATION, RIGHT-HANDED BETA-HELIX, TMAFP
1102	1ezg	A	276	356	6.4e-09	1.21	-0.09		THERMAL HYSTERESIS PROTEIN ISOFORM YL-1; CHAIN: A, B;	ANTIFREEZE PROTEIN INSECT ANTIFREEZE PROTEIN, THERMAL HYSTERESIS, TENEBRIO 2 MOLITOR, IODINATION, RIGHT-HANDED BETA-HELIX, TMAFP

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1102	1ezg	A	289	368	3.2e-09	1.12	-0.11		THERMAL HYSTERESIS PROTEIN ISOFORM YL-1; CHAIN: A, B;	ANTIFREEZE PROTEIN INSECT ANTIFREEZE PROTEIN, THERMAL HYSTERESIS, TENEBRIO 2 MOLITOR, IODINATION, RIGHT-HANDED BETA-HELIX, TMAFP
1102	1ezg	A	312	392	1.4e-10	1.07	-0.02		THERMAL HYSTERESIS PROTEIN ISOFORM YL-1; CHAIN: A, B;	ANTIFREEZE PROTEIN INSECT ANTIFREEZE PROTEIN, THERMAL HYSTERESIS, TENEBRIO 2 MOLITOR, IODINATION, RIGHT-HANDED BETA-HELIX, TMAFP
1102	1ezg	A	338	420	4.8e-12	0.70	-0.11		THERMAL HYSTERESIS PROTEIN ISOFORM YL-1; CHAIN: A, B;	ANTIFREEZE PROTEIN INSECT ANTIFREEZE PROTEIN, THERMAL HYSTERESIS, TENEBRIO 2 MOLITOR, IODINATION, RIGHT-HANDED BETA-HELIX, TMAFP
1102	1ezg	A	359	440	4.8e-08	0.20	-0.17		THERMAL HYSTERESIS PROTEIN ISOFORM YL-1; CHAIN: A, B;	ANTIFREEZE PROTEIN INSECT ANTIFREEZE PROTEIN, THERMAL HYSTERESIS, TENEBRIO 2 MOLITOR, IODINATION, RIGHT-HANDED BETA-HELIX, TMAFP
1102	1klo		146	301	3.2e-08	0.02	-0.19		LAMININ; CHAIN: NULL;	GLYCOPROTEIN
1102	1klo		168	342	1.6e-10	0.20	-0.19		LAMININ; CHAIN: NULL;	GLYCOPROTEIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1102	1klo		240	419	3.2e-09	0.02	-0.19		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
1102	4m2		331	394	4.8e-06	-0.12	0.06		METALLOTHIONEIN METALLOTHIONEIN ISOFORM II 4MT2 3	
1102	9wga	A	126	266	3.2e-09	0.16	-0.14		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
1102	9wga	A	142	298	6.4e-12	0.15	-0.18		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
1102	9wga	A	193	370	8e-11	0.04	-0.11		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
1102	9wga	A	276	438	6.4e-09	0.06	-0.15		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
1107	1qo7	A	89	187	0.0096	0.22	0.34		EPOXIDE HYDROLASE; CHAIN: A, B;	EPOXIDE HYDROLASE EH; EPOXIDE HYDROLASE, ALPHA/BETA HYDROLASE
1109	1abt	A	34	111	0.0014	0.16	0.23		TOXIN ALPHA-BUNGAROTOXIN COMPLEXED WITH THE 185 - 196 FRAGMENT OF 1ABT 3 THE ALPHA-SUBUNIT OF THE TORPEDO NICOTINIC	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1109	1lci	A	34	111	0.0029	-0.03	0.04		ACETYLCHOLINE 1ABT 4 RECEPTOR (NMR, 4 STRUCTURES) 1ABT 5	
									CARDIOTOXIN V; CHAIN: A, B;	CYTOTOXIN CTX A5; VENOM, CYTOTOXIN, CARDIOTOXIN, MULTIGENE FAMILY, SIGNAL
1110	1ajj		197	229	9.6e-10	0.35	0.84		LOW-DENSITY LIPOPROTEIN RECEPTOR; CHAIN: NULL;	RECEPTOR LR5; RECEPTOR, LDL RECEPTOR, CYSTEINE-RICH MODULE, CALCIUM
1110	1ajj		197	233	9.6e-09	0.04	-0.03		LOW-DENSITY LIPOPROTEIN RECEPTOR; CHAIN: NULL;	RECEPTOR LR5; RECEPTOR, LDL RECEPTOR, CYSTEINE-RICH MODULE, CALCIUM
1110	1cr8	A	195	240	4.8e-09	0.15	0.01		LOW-DENSITY LIPOPROTEIN RECEPTOR RELATED PROTEIN; CHAIN: A;	LIPID BINDING PROTEIN RECEPTOR, LIGAND BINDING, CALCIUM BINDING, LDLR, LRP, LIPID 2 BINDING PROTEIN
1110	1d2j	A	197	229	1.4e-09	0.43	0.49		LOW-DENSITY LIPOPROTEIN RECEPTOR; CHAIN: A;	SIGNALING PROTEIN LR6*, RECEPTOR, LDLR, CYSTEINE-RICH MODULE, CALCIUM
									LIGAND-2 BINDING, FAMILIAL HYPERCHOLESTEROLEMIA	
1110	1d2l	A	193	229	2.9e-10	1.00	0.49		LIPOPROTEIN RECEPTOR RELATED PROTEIN; CHAIN: A;	SIGNALING PROTEIN LIGAND BINDING, CALCIUM BINDING, COMPLEMENT-LIKE REPEAT, 2 RECEPTOR, SIGNALING
										PROTEIN
1110	1t5y	A	157	233	3.2e-09	0.41	-0.14		LOW-DENSITY LIPOPROTEIN	LIPID BINDING PROTEIN LDL

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									RECEPTOR; CHAIN: A;	RECEPTOR; BETA HAIRPIN, 3-10 HELIX, CALCIUM BINDING
1110	1f5y	A	190	254	4.8e-10	0.57	-0.09		LOW-DENSITY LIPOPROTEIN RECEPTOR; CHAIN: A;	LIPID BINDING PROTEIN LDL RECEPTOR; BETA HAIRPIN, 3-10 HELIX, CALCIUM BINDING
1110	1f5y	A	88	157	6.4e-09	0.13	-0.20		LOW-DENSITY LIPOPROTEIN RECEPTOR; CHAIN: A;	LIPID BINDING PROTEIN LDL RECEPTOR; BETA HAIRPIN, 3-10 HELIX, CALCIUM BINDING
1110	1ldl		195	229	4.8e-10	0.52	0.30		LOW-DENSITY LIPOPROTEIN RECEPTOR; 1LDL 4 CHAIN: NULL; 1LDL 5	BINDING PROTEIN LB1; 1LDL 7 LDL RECEPTOR CYSTEINE-RICH REPEAT 1LDL 15
1110	9wga	A	117	274	3.2e-18			54.99	LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
1110	9wga	A	40	209	1.6e-13	0.23	-0.19		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
1110	9wga	A	90	238	3.2e-18	0.12	-0.18		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
1118	1c0t	A	56	175	3.2e-38	-0.01	0.59		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1118	1c0t	B	47	159	3.2e-38	-0.26	0.31		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN);	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1118	1c1c	B	47	175	4.8e-42	-0.47	0.54		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1118	1c9t	A	56	175	4.8e-38	-0.14	0.58		HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5'-CHAIN: T; DNA (5'-CHAIN: P;	TRANSFERASE/IMMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184LE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/IMMUNE 3 SYSTEM/DNA
1118	1c9t	B	47	175	1.6e-40	-0.26	0.81		HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5'-CHAIN: T; DNA (5'-CHAIN: P;	TRANSFERASE/IMMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184LE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/IMMUNE 3 SYSTEM/DNA
1118	1har		2	179	9.6e-37			50.82	REVERSE TRANSCRIPTASE HIV-1 REVERSE	



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									TRANSCRIPTASE (AMINO-TERMINAL HALF) (FINGERS 1HAR 3 AND PALM SUBDOMAINS) (RT16) (E.C.2.7.7.49) 1HAR 4	
1118	1rth	A	47	175	1.6e-42	-0.22	0.84		HIV-1 REVERSE TRANSCRIPTASE; 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1118	1rth	B	47	175	4.8e-43	-0.21	0.54		HIV-1 REVERSE TRANSCRIPTASE; 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1118	1vrt	A	47	175	1.6e-42	-0.34	0.81		HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
1118	1vrt	B	47	175	1.6e-42	-0.14	0.57		HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
1118	3hvt	B	47	175	3.2e-42	-0.04	0.24		NUCLEOTIDYLTRANSFERASE REVERSE TRANSCRIPTASE (E.C.2.7.7.49) 3HVT 3	
1119	1clg	A	64	138	2.4e-09	1.36	-0.18		TROPOMYOSIN; CHAIN: A, B, C, D	CONTRACTILE PROTEIN TROPOMYOSIN COILED-COIL ALPHA-HELICAL,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verity score	PME score	SEQFOLD score	Compound	PDB annotation
1119	1c1g	A	64	156	9.6e-09	0.91	-0.19		TROPOMYOSIN; CHAIN: A, B, C, D	CONTRACTILE PROTEIN CONTRACTILE PROTEIN TROPOMYOSIN COILED-COIL ALPHA-HELICAL, CONTRACTILE PROTEIN
1119	1c1i		66	152	4.3e-08	0.70	-0.20		COLICIN 1A; CHAIN: NULL;	TRANSMEMBRANE PROTEIN COLICIN, BACTERIOCIN, ION CHANNEL FORMATION, TRANSMEMBRANE 2 PROTEIN
1119	1e23	A	64	157	4.8e-15	0.95	-0.20		SYNTAXIN-1A; CHAIN: A, B, C;	ENDOCYTOSIS/EXOCYTOSIS SYNAPTOTAGMIN ASSOCIATED 35 KDA PROTEIN, P35A, THREE HELIX BUNDLE
1119	1q1q	A	71	135	2.9e-09	0.13	-0.20		GLUTAMINYL-TRNA SYNTHETASE; CHAIN: A; TRNA GLN II; CHAIN: B;	COMPLEX (TRNA SYNTHETASE/TRNA) GLNRS; TRNA SYNTHETASE, GLUTAMINE, TRNAGLN, E. COLI, COMPLEX
1119	1quu	A	64	152	1.9e-12	0.59	-0.20		HUMAN SKELETAL MUSCLE ALPHA-ACTININ 2; CHAIN: A;	CONTRACTILE PROTEIN TRIPLE-HELIX COILED COIL, CONTRACTILE PROTEIN
1119	1req	A	64	152	1.4e-11	0.77	-0.19		METHYLMALONYL-COA MUTASE; CHAIN: A, B, C, D;	ISOMERASE ISOMERASE, MUTASE, INTRAMOLECULAR TRANSFERASE
1125	1ak2		31	117	9.6e-22	-0.44	0.41		ADENYLYLATE KINASE ISOENZYME-2; CHAIN: NULL;	PHOSPHOTRANSFERASE ATP:AMP PHOSPHOTRANSFERASE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1125	1aky		31	117	1.1e-19	-0.23	0.11		ADENYLATE KINASE; IAKY 4 CHAIN: NULL; IAKY 5	MYOKINASE, NUCLEOSIDE MONOPHOSPHATE KINASE, PHOSPHOTRANSFERASE
1125	1edv	A	31	116	1.6e-20	-0.55	0.00		ADENYLATE KINASE; CHAIN: A;	TRANSFERASE(PHOSPHOTRANSFERASE)
1125	1qf9	A	44	117	6.4e-23	-0.10	0.41		URIDYLMONOPHOSPHATE/CYTIDYLMONOPHOSPHATE KINASE; CHAIN: A;	KINASE UMP/CMP KINASE; NUCLEOSIDE MONOPHOSPHATE KINASE, NMP KINASE, PHOSPHORYL 2 TRANSFER, TRANSITION STATE ANALOG COMPLEX, TRANSFERASE
1125	1ukz		43	117	1.6e-23	-0.26	0.99		TRANSFERASE URIDYLATE KINASE (B.C.2.7.4.-) COMPLEXED WITH ADP AND AMP 1UKZ 3	
1125	1zak	A	46	117	9.6e-15	-0.45	0.05		ADENYLATE KINASE; CHAIN: A, B;	TRANSFERASE ATP-AMP-PHOSPHOTRANSFERASE, TRANSFERASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1125	1zim		31	117	9.6e-21	-0.57	0.15		ADENYLATE KINASE; CHAIN: NULL;	PHOSPHOTRANSFERASE ADK; PHOSPHOTRANSFERASE, ZINC FINGER
1125	2ak3	A	31	119	1.6e-19	-0.17	0.05		TRANSFERASE (PHOSPHOTRANSFERASE) ADENYLATE KINASE ISOENZYME-3 (GTP: AMP PHOSPHOTRANSFERASE) 2AK3 3 (E.C.2.7.4.10) 2AK3 4	
1125	3adk		43	117	3.2e-26	-0.04	0.75		TRANSFERASE(PHOSPHOTRANSFERASE) ADENYLATE KINASE (E.C.2.7.4.3) 3ADK 4	
1135	1q92	A	7	42	0.0048	-0.79	0.11		THIOREDOXIN PEROXIDASE 2; CHAIN: A, B;	OXIDOREDUCTASE HEME-BINDING PROTEIN 23 KD, HBP23; THIOREDOXIN FOLD, OXIDOREDUCTASE
1162	1mcy	G	91	117	1.6e-11	0.01	-0.20		DNA; CHAIN: A, B, D, E; CONSENSUS ZINC FINGER PROTEIN; CHAIN: C, F, G;	COMPLEX (ZINC FINGER/DNA) ZINC FINGER, PROTEIN-DNA INTERACTION, PROTEIN DESIGN, 2 CRYSTAL STRUCTURE, COMPLEX (ZINC FINGER/DNA)
1165	1erj	A	35	114	0.0024	0.55	0.80		TRANSCRIPTIONAL REPRESSOR TUP1; CHAIN: A, B, C;	TRANSCRIPTION INHIBITOR BETA-PROPELLER

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1165	1got	B	33	127	0.0096	0.21	0.03		GT-ALPHA/GI-ALPHA CHIMERA; CHAIN: A; GT-BETA; CHAIN: B; GT-GAMMA; CHAIN: G;	COMPLEX (GTP-BINDING/TRANSDUCER) BETA1, TRANSDUCIN BETA SUBUNIT; GAMMA1, TRANSDUCIN GAMMA SUBUNIT; COMPLEX (GTP-BINDING/TRANSDUCER), G PROTEIN, HETEROTRIMER 2 SIGNAL TRANSDUCTION
1165	1got	B	35	110	0.00024	0.53	0.90		GT-ALPHA/GI-ALPHA CHIMERA; CHAIN: A; GT-BETA; CHAIN: B; GT-GAMMA; CHAIN: G;	COMPLEX (GTP-BINDING/TRANSDUCER) BETA1, TRANSDUCIN BETA SUBUNIT; GAMMA1, TRANSDUCIN GAMMA SUBUNIT; COMPLEX (GTP-BINDING/TRANSDUCER), G PROTEIN, HETEROTRIMER 2 SIGNAL TRANSDUCTION
1193	1aln	A	29	226	1.6e-91	0.58	1.00		B*3501; CHAIN: A, B; PEPTIDE VPLRPMTY; CHAIN: C;	COMPLEX (ANTIGEN/PEPTIDE) B35; MAJOR HISTOCOMPATIBILITY ANTIGEN, MHC, HLA, HLA-B3501, HIV, 2 NEF, COMPLEX (ANTIGEN/PEPTIDE)
1193	1aln	A	29	243	1.6e-91			55.07	B*3501; CHAIN: A, B; PEPTIDE VPLRPMTY; CHAIN: C;	COMPLEX (ANTIGEN/PEPTIDE) B35; MAJOR HISTOCOMPATIBILITY ANTIGEN, MHC, HLA, HLA-B3501, HIV, 2 NEF, COMPLEX

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1193	1a6a	B	134	220	0.0014	-0.21	0.11		HLA-DR3; CHAIN: A, B; CLIP; CHAIN: C;	(ANTIGEN/PEPTIDE) COMPLEX (TRANSMEMBRANE/GLYCOPROTEIN) MHC GLYCOPROTEIN, COMPLEX (TRANSMEMBRANE/GLYCOPROTEIN)
1193	1a6z	A	29	245	3.2e-64			75.13	HFE; CHAIN: A, C; BETA-2-MICROGLOBULIN; CHAIN: B, D	MHC CLASS I COMPLEX HFE, HEREDITARY HEMOCHROMATOSIS, MHC CLASS I
1193	1agd	A	29	226	9.6e-92	0.57	1.00		B*0801; CHAIN: A; BETA-2-MICROGLOBULIN; CHAIN: B; HIV-1 GAG PEPTIDE (GGKKKKYKL - INDEX PEPTIDE); CHAIN: C;	HISTOCOMPATIBILITY COMPLEX B8; B2M; PEPTIDE HLA B8, HIV, MHC CLASS I, HISTOCOMPATIBILITY COMPLEX
1193	1agd	A	29	243	9.6e-92			61.21	B*0801; CHAIN: A; BETA-2-MICROGLOBULIN; CHAIN: B; HIV-1 GAG PEPTIDE (GGKKKKYKL - INDEX PEPTIDE); CHAIN: C;	HISTOCOMPATIBILITY COMPLEX B8; B2M; PEPTIDE HLA B8, HIV, MHC CLASS I, HISTOCOMPATIBILITY COMPLEX
1193	1agq	B	121	220	0.00032	-0.04	0.06		HLA-DRI CLASS II HISTOCOMPATIBILITY PROTEIN; CHAIN: A, B, D, E, G, H, J, K; HLA-A2; CHAIN: C, F, I, L;	COMPLEX (MHC PROTEIN/ANTIGEN) DRA, DRB1 01010; COMPLEX (MHC PROTEIN/ANTIGEN), HISTOCOMPATIBILITY ANTIGEN
1193	1bx2	B	134	220	0.00048	-0.15	0.06		HLA-DR2; CHAIN: A, D; HLA-	IMMUNE SYSTEM HLA-DR2,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									DR2; CHAIN: B, E; HLA-DR2; CHAIN: C, F;	MYELIN BASIC PROTEIN, MULTIPLE SCLEROSIS, 2 AUTOIMMUNITY, IMMUNE SYSTEM
1193	1cd1	A	53	217	1.6e-09	-0.31	0.11		CD1; CHAIN: A, B, C, D;	CD1 MCD1D.1; CD1, IMMUNOLOGY, MHC, TCR, GLYCOPROTEIN, SIGNAL, 2 IMMUNOGLOBULIN FOLD, T-CELL
1193	1duz	A	29	226	6.4e-91	0.46	1.00		HLA-A*0201; CHAIN: A, D; BETA-2 MICROGLOBULIN; CHAIN: B, E; HTLV-1 OCTAMERIC TAX PEPTIDE; CHAIN: C, F;	IMMUNE SYSTEM IMMUNOGLOBULIN FOLD
1193	1efk	A	29	226	1.6e-91	0.49	1.00		HLA-CW3 (HEAVY CHAIN); CHAIN: A; BETA-2-MICROGLOBULIN; CHAIN: B; PEPTIDE FROM IMPORTIN ALPHA-2; CHAIN: C; NATURAL KILLER CELL RECEPTOR KIR2DL2; CHAIN: D, E;	IMMUNE SYSTEM MHC, HLA, CLASS I, KIR, NK CELL RECEPTOR, IMMUNOGLOBULIN 2 FOLD, RECEPTOR/MHC COMPLEX
1193	1hoc	A	29	245	8e-87			64.64	HISTOCOMPATIBILITY ANTIGEN MURINE CLASS I MAJOR HISTOCOMPATIBILITY COMPLEX CONSISTING 1HOC 3 OF H-2D=B=, B2-	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									MICROGLOBULIN, AND A 9-RESIDUE PEPTIDE IHOC 4	
1193	1hsa	A	29	226	1.6e-91	0.42	1.00		HISTOCOMPATIBILITY ANTIGEN HUMAN CLASS I HISTOCOMPATIBILITY ANTIGEN I HSA 3 /HLA-B*ASTERISK)2705\$ I HSA 4	
1193	1hsa	A	29	243	1.6e-91			52.86	HISTOCOMPATIBILITY ANTIGEN HUMAN CLASS I HISTOCOMPATIBILITY ANTIGEN I HSA 3 /HLA-B*ASTERISK)2705\$ I HSA 4	
									HISTOCOMPATIBILITY ANTIGEN CLASS I HISTOCOMPATIBILITY ANTIGEN I HSA 3 /HLA-B*ASTERISK)2705\$ I HSA 4	
1193	1hsb	A	29	226	1.6e-91	0.50	1.00		HISTOCOMPATIBILITY ANTIGEN CLASS I HISTOCOMPATIBILITY ANTIGEN AW68.1 (LEUCOCYTE I HSB 3 ANTIGEN) I HSB 4	
									HISTOCOMPATIBILITY ANTIGEN CLASS I HISTOCOMPATIBILITY ANTIGEN AW68.1 (LEUCOCYTE I HSB 3 ANTIGEN) I HSB 4	
1193	1hsb	A	29	245	1.6e-91			59.60	HISTOCOMPATIBILITY ANTIGEN CLASS I HISTOCOMPATIBILITY ANTIGEN AW68.1 (LEUCOCYTE I HSB 3 ANTIGEN) I HSB 4	
									HISTOCOMPATIBILITY ANTIGEN CLASS I HISTOCOMPATIBILITY ANTIGEN AW68.1 (LEUCOCYTE I HSB 3 ANTIGEN) I HSB 4	
1193	1hd9	A	29	226	1.6e-89	0.40	1.00		MHC CLASS I H-2LD HEAVY CHAIN; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; NANO-PEPTIDE; CHAIN: C;	MAJOR HISTOCOMPATIBILITY COMPLEX LD; MAJOR HISTOCOMPATIBILITY COMPLEX, LD



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1193	1ld9	A	29	246	1.6e-89			68.80	MHC CLASS I H-2LD HEAVY CHAIN; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; NANO-PEPTIDE; CHAIN: C;	MAJOR HISTOCOMPATIBILITY COMPLEX LD; MAJOR HISTOCOMPATIBILITY COMPLEX, LD
1193	1mh	A	30	246	4.8e-86			52.55	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN HLA-E; CHAIN: A, C; BETA-2-MICROGLOBULIN; CHAIN: B, D; PEPTIDE (VMAPTVLL); CHAIN: F, Q;	MAJOR HISTOCOMPATIBILITY COMPLEX MHC NONCLASSICAL CHAIN, MHC-E, HLA-E, MHC CLASS HLA-E, HLA E, MAJOR HISTOCOMPATIBILITY COMPLEX, MHC, HLA, 2 BETA 2 MICROGLOBULIN, PEPTIDE, LEADER PEPTIDE, 3 NON-CLASSICAL MHC, CLASS IB MHC
1193	1os2	A	29	226	1.6e-87	0.66	1.00		MHC CLASS I H-2KB HEAVY CHAIN; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; VESICULAR STOMATITIS VIRUS NUCLEOPROTEIN; CHAIN: C;	COMPLEX (MHC I/PEPTIDE) VSV-8; MHC/PEPTIDE COMPLEX, TRANSMEMBRANE PROTEIN, THYMIC 2 SELECTION, COMPLEX (MHC I/PEPTIDE)
1193	1os2	A	29	245	1.6e-87			60.44	MHC CLASS I H-2KB HEAVY CHAIN; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; VESICULAR STOMATITIS VIRUS NUCLEOPROTEIN; CHAIN: C;	COMPLEX (MHC I/PEPTIDE) VSV-8; MHC/PEPTIDE COMPLEX, TRANSMEMBRANE PROTEIN, THYMIC 2 SELECTION, COMPLEX (MHC I/PEPTIDE)
1193	1qo3	A	30	226	8e-90	0.60	1.00		MHC CLASS I H-2DD HEAVY	COMPLEX (NK RECEPTOR/MHC

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsI Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN: CHAIN: A; BETA-2-MICROGLOBULIN; CHAIN: B; HIV ENVELOPE GLYCOPROTEIN 120 PEPTIDE; CHAIN: P; LY49A; CHAIN: C; D;	CLASS II H-2 CLASS I HISTOCOMPATIBILITY ANTIGEN, B2M; NK-CELL SURFACE GLYCOPROTEIN YE1/48, NK CELL, INHIBITORY RECEPTOR, MHC-I, C-TYPE LECTIN-LIKE, 2 HISTOCOMPATIBILITY, B2M, LY49, LY-49
1193	1qgd	A	30	226	1.6e-90	0.69	1.00		HISTOCOMPATIBILITY LEUKOCYTE ANTIGEN (HLA)-CW4 CHAIN: A; BETA-2-MICROGLOBULIN; CHAIN: B; HLA-CW4 SPECIFIC PEPTIDE; CHAIN: C;	IMMUNE SYSTEM IMMUNOGLOBULIN (IG)-LIKE DOMAIN, ALPHA HELIX, BETA SHEET, 2 IMMUNE SYSTEM
1193	1tmc	A	29	200	1.6e-79			76.13	HISTOCOMPATIBILITY ANTIGEN TRUNCATED HUMAN CLASS I HISTOCOMPATIBILITY ANTIGEN HLA-AW68 1TMC3 COMPLEXED WITH A DECAMERIC PEPTIDE (EVAPPEYHRK) 1TMC4	
1193	1zag	A	28	244	1.6e-57			67.63	ZINC-ALPHA-2-GLYCOPROTEIN; CHAIN: A, B, C, D;	LIPID MOBILIZATION FACTOR ZN-ALPHA-2-GLYCOPROTEIN, ZAG LIPID MOBILIZATION FACTOR, SECRETED MHC CLASS I HOMOLOG

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1193	3ftu	A	27	246	3.2e-45			56.39	NEONATAL FC RECEPTOR; CHAIN: A, C, E; BETA-2-MICROGLOBULIN; CHAIN: B, D, F;	FCRN, BRAMBELL RECEPTOR; COMPLEX (IMMUNOGLOBULIN/BINDING PROTEIN)
1195	1d1d	A	69	167	3.2e-29	-0.48	0.10		CAPSID PROTEIN; CHAIN: A;	VIRUS/VIRAL PROTEIN TWO INDEPENDENT DOMAINS HELICAL BUNDLES, VIRUS/VIRAL PROTEIN
1195	1em9	A	57	167	6.4e-30	-0.16	0.33		GAG POLYPROTEIN CAPSID PROTEIN P27; CHAIN: A, B;	VIRUS/VIRAL PROTEIN
1195	1em9	A	84	172	1.4e-15	-0.15	0.37		GAG POLYPROTEIN CAPSID PROTEIN P27; CHAIN: A, B;	VIRUS/VIRAL PROTEIN
1195	1em9	B	67	167	6.4e-28	-0.18	0.33		GAG POLYPROTEIN CAPSID PROTEIN P27; CHAIN: A, B;	VIRUS/VIRAL PROTEIN
1195	1qfj	B	86	139	9.6e-07	0.27	0.74		HIS TAG; CHAIN: A; HTLV-I CAPSID PROTEIN; CHAIN: B;	VIRUS/VIRAL PROTEIN HTLV-I, CAPSID PROTEIN, RETROVIRUS, TWO-DOMAIN PROTEIN, 2 ALPHA HELICAL PROTEIN, HETERONUCLEAR NMR SPECTROSCOPY, 3 VIRUS/VIRAL PROTEIN
1212	1ael	A	26	111	8e-17	0.43	0.69		TROPINONE REDUCTASE-I; CHAIN: A, B;	OXIDOREDUCTASE TROPANE ALKALOID BIOSYNTHESIS, REDUCTION OF 2 TROPINONE TO TROPINE, SHORT-CHAIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1212	lael	B	26	111	8e-17	0.41	0.63		TROPINONE REDUCTASE-I; CHAIN: A, B;	DEHYDROGENASE OXIDOREDUCTASE OXIDOREDUCTASE, TROPANE ALKALOID BIOSYNTHESIS, REDUCTION OF 2 TROPINONE TO TROPINE, SHORT-CHAIN DEHYDROGENASE
1212	lbdb		27	111	1.3e-20	0.02	0.21		CIS-BIPHENYL-2,3-DIHYDRODIOL-2,3-DEHYDROGENASE; CHAIN: NULL;	OXIDOREDUCTASE NAD-DEPENDENT OXIDOREDUCTASE, SHORT-CHAIN ALCOHOL 2 DEHYDROGENASE, PCB DEGRADATION
1212	lcyd	A	26	109	3.2e-12	0.56	0.43		CARBONYL REDUCTASE; CHAIN: A, B, C, D;	OXIDOREDUCTASE SHORT-CHAIN DEHYDROGENASE, OXIDOREDUCTASE
1212	lfmc	A	21	107	6.4e-24	0.24	1.00		7 ALPHA-HYDROXYSTEROID DEHYDROGENASE; CHAIN: A, B;	OXIDOREDUCTASE SHORT-CHAIN DEHYDROGENASE/REDUCTASE, BILE ACID CATABOLISM
1212	lhdc	A	27	115	1.6e-20	0.53	0.29		OXIDOREDUCTASE 3-ALPHA, 20-BETA-HYDROXYSTEROID DEHYDROGENASE (E.C.1.1.1.53) 1HDC3 COMPLEXED WITH CARBENOXOLONE 1HDC4	
1212	loaa		27	106	9.6e-10	0.19	0.77		SEPIAPTERIN REDUCTASE; CHAIN: NULL;	OXIDOREDUCTASE SEPIAPTERIN REDUCTASE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1212	1ybv	A	25	107	6.4e-22	0.64	0.90		TRIHYDROXYNAPHTHALENE REDUCTASE; CHAIN: A, B;	TETRAHYDROBIPTERIN, OXIDOREDUCTASE
1212	2ae2	A	26	111	1.1e-15	0.61	0.69		TROPINONE REDUCTASE-II; CHAIN: A, B;	OXIDOREDUCTASE, TROPANE ALKALOID BIOSYNTHESIS, REDUCTION OF 2 TROPINONE TO PSEUDOTROPINE, SHORT-CHAIN DEHYDROGENASE
1213	1mgs	A	25	57	8e-12	-0.85	0.42		CHEMOKINE(GROWTH FACTOR) HUMAN MELANOMA GROWTH STIMULATING ACTIVITY (MGSA/GRO ALPHA) IMGs 3 (NMR, 25 STRUCTURES) IMGs 4	
1213	1mi2	A	25	56	9.6e-12	-0.82	0.01		MACROPHAGE INFLAMMATORY PROTEIN-2; CHAIN: A, B;	CYTOKINE MIP-2, CHEMOKINE, NMR, CYTOKINE
1213	1pfm	A	27	57	1.6e-11	-0.78	0.93		PF4-M2 CHIMERA; 1PFM 7 CHAIN: A, B, C, D; 1PFM 8	CYTOKINE PLATELET FACTOR 4, PLATELET FACTOR M2; 1PFM 9
1213	1plf	A	27	57	4.8e-10	-0.46	0.96		PLATELET FACTOR PLATELET FACTOR 4 1PLF 3	
1213	1puk	A	25	57	9.6e-12	-0.85	0.28		GROB[5-73]; CHAIN: A,B;	CHEMOKINE CHEMOKINE 15-O,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1213	1thp	A	27	57	1.6e-11	-0.46	0.98		PLATELET FACTOR 4 (HPF4) (HUMAN RECOMBINANT) 1RHP 3	HUMAN CHEMOKINE GROB[5-73], CXC CHEMOKINE
1213	1tvx	A	25	55	1.3e-11	-0.72	0.37		NEUTROPHIL ACTIVATING PEPTIDE 2 VARIANT; CHAIN: A, B, C, D;	CYTOKINE NAP-2; CYTOKINE
1213	1tvx	B	25	55	1.3e-11	-0.84	0.28		NEUTROPHIL ACTIVATING PEPTIDE 2 VARIANT; CHAIN: A, B, C, D;	CYTOKINE NAP-2; CYTOKINE
1224	1a5r		64	107	3.2e-18	-0.43	0.10		SUMO-1; CHAIN: NULL;	TARGETING PROTEIN PIC1, GMP1, UBL1, SENTRIN; SUMO-1, POST-TRANSLATIONAL PROTEIN MODIFICATION, 2 UBIQUITIN-LIKE PROTEINS, TARGETING PROTEIN
1224	1euv	B	63	107	3.2e-17	0.17	0.46		ULP1 PROTEASE; CHAIN: A; UBIQUITIN-LIKE PROTEIN SMT3; CHAIN: B;	HYDROLASE SUMO HYDROLASE, UBIQUITIN-LIKE PROTEASE 1, SMT3 HYDROLASE 2 DESUMOYLATING ENZYME, CYSTEINE PROTEASE, SUMO PROCESSING 3 ENZYME, SMT3 PROCESSING ENZYME, NABH4, THIOHEMACETAL, 4

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1231	2di	A	68	108	0.0083	-0.47	0.13		MHC CLASS I NK CELL RECEPTOR PRECURSOR, CHAIN: A;	IMMUNE SYSTEM P58 NATURAL KILLER CELL RECEPTOR; KIR, NATURAL KILLER RECEPTOR, INHIBITORY RECEPTOR, 2 IMMUNOGLOBULIN
1240	1a3r	L	25	248	4.8e-65			82.35	IGG2A; CHAIN: L, H; HUMAN RHINOVIRUS CAPSID PROTEIN VP2; CHAIN: P;	COMPLEX (IMMUNOGLOBULIN/VIRAL PEPTIDE) ANTIBODY 8F5; IMMUNOGLOBULIN, ANTIBODY, RHINOVIRUS, NEUTRALIZATION, 2 CONTINUOUS EPTOPE, COMPLEX (IMMUNOGLOBULIN/VIRAL PEPTIDE)
1240	1a4j	L	25	246	1.6e-62			83.80	IMMUNOGLOBULIN, DIELS ALDER CATALYTIC ANTIBODY; CHAIN: L, H, A, B;	IMMUNOGLOBULIN IMMUNOGLOBULIN, ANTIBODY, CATALYTIC ANTIBODY, DIELS ALDER, 2 GERM LINE
1240	1ad9	L	27	248	9.6e-64			83.32	FAB FRAGMENT CTM01; CHAIN: L, H, A, B;	IMMUNOGLOBULIN IMMUNOGLOBULIN, FAB FRAGMENT
1240	1ae6	L	25	248	9.6e-64			85.78	ANTIBODY CTM01; CHAIN: L;	IMMUNOGLOBULIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1240	1aif	L	25	248	3.2e-63			81.78	ANTI-IDIOTYPIC FAB 409.5.3 (IGG2A) FAB; CHAIN: A, B, L, H	IMMUNOGLOBULIN, FAB FRAGMENT, HUMANISATION
1240	1b4j	L	27	248	3.2e-62			83.18	ANTIBODY; CHAIN: L, H;	ANTIBODY ENGINEERING
1240	1b6d	A	27	246	3.2e-62			82.05	IMMUNOGLOBULIN; CHAIN: A, B;	ANTIBODY ENGINEERING, HUMANIZED AND CHIMERIC ANTIBODIES, 2 FAB, X-RAY STRUCTURES, GAMMA-INTERFERON
1240	1baf	L	28	248	1.6e-65	-0.35	0.27		IMMUNOGLOBULIN FAB FRAGMENT OF MURINE MONOCLONAL ANTIBODY AN02 COMPLEX IBAF 3 WITH ITS HAPTEN (2,2,6,6-TETRAMETHYL-1-PIPERIDINYLOXY-IBAF 4 DINITROPHENYL) IBAF 5	IMMUNOGLOBULIN, KAPPA LIGHT-CHAIN DIMER HEADER
1240	1b6j	L	25	244	6.4e-61			86.24	IMMUNOGLOBULIN FAB FRAGMENT OF MONOCLONAL ANTIBODY B72.3 IBB1 3 (MURINE/HUMAN CHIMERA) IBB1 4	



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Pst Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1240	1bj1	L	27	247	6.4e-64			83.21	FAB FRAGMENT; CHAIN: L, H, I, K; VASCULAR ENDOTHELIAL GROWTH FACTOR; CHAIN: V, W;	COMPLEX (ANTIBODY/ANTIGEN) FAB-12; VEGF; COMPLEX (ANTIBODY/ANTIGEN), ANGIOGENIC FACTOR
1240	1bz7	A	25	240	4.8e-57			86.65	ANTIBODY R24 (LIGHT CHAIN); CHAIN: A; ANTIBODY R24 (HEAVY CHAIN); CHAIN: B;	IMMUNE SYSTEM ANTIBODY (FAB FRAGMENT), IMMUNE SYSTEM
1240	1cl2	A	25	248	8e-62			82.69	ANTIBODY FRAGMENT FAB; CHAIN: A; ANTIBODY FRAGMENT FAB; CHAIN: B;	IMMUNE SYSTEM ANTIBODY-ANTIGEN COMPLEX, SCFV FRAGMENT, CDRH3, MUSK 2 ODORANT, ODORANT SPECIFICITY, IMMUNE SYSTEM
1240	1cf8	L	28	248	1.6e-65	-0.18	0.16		CATALYTIC ANTIBODY 19A4 (LIGHT CHAIN); CHAIN: L; CATALYTIC ANTIBODY 19A4 (HEAVY CHAIN); CHAIN: H;	CATALYTIC ANTIBODY CATALYTIC ANTIBODY, TERPENOID SYNTHASE, CARBOCATION, 2 CYCLIZATION CASCADE
1240	1cly	L	28	248	4.8e-63			83.48	IGG FAB (HUMAN IGG1, KAPPA); CHAIN: L, H;	IMMUNOGLOBULIN CBR96 FAB (IMMUNOGLOBULIN); IMMUNOGLOBULIN, IMMUNOGLOBULIN C REGION, GLYCOPROTEIN, ANTIB
1240	1cvs	C	87	231	9.6e-08	-0.29	0.12		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN:	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									C, D;	DIMERIZATION, GROWTH FACTOR/GROWTH FACTOR RECEPTOR
1240	1cvs	D	27	231	9.6e-06	0.04	0.45		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH FACTOR/GROWTH FACTOR RECEPTOR
1240	1d5m	B	42	245	1.6e-44	-0.35	0.12		HLA CLASS II HISTOCOMPATIBILITY ANTIGEN; CHAIN: A; HLA CLASS II HISTOCOMPATIBILITY ANTIGEN; CHAIN: B; ENTEROTOXIN TYPE B; CHAIN: C; PEPTIDE INHIBITOR; CHAIN: D;	IMMUNE SYSTEM HLA-DR4; HLA-DR4; SEB, SUPERANTIGEN; COMPLEX (MHC CLASS II/SUPERANTIGEN), IMMUNE SYSTEM
1240	1epf	A	36	231	9.6e-07	-0.22	0.07		NEURAL CELL ADHESION MOLECULE; CHAIN: A, B, C, D;	CELL ADHESION NCAM; NCAM, IMMUNOGLOBULIN FOLD, GLYCOPROTEIN
1240	1f11	A	28	246	1.6e-65	-0.05	0.53		F124 IMMUNOGLOBULIN (KAPPA LIGHT CHAIN); CHAIN: A, C, F124 IMMUNOGLOBULIN (IGG1 HEAVY CHAIN); CHAIN: B, D;	IMMUNE SYSTEM IMMUNOGLOBULIN, ANTIBODY, FAB, HEPATITIS B, PRES2
1240	1f5w	A	40	139	9.6e-09	0.47	0.94		COXSACKIE VIRUS AND	VIRUS/VIRAL PROTEIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									ADENOVIRUS RECEPTOR; CHAIN: A, B;	RECEPTOR IMMUNOGLOBULIN V DOMAIN FOLD, SYMMETRIC DIMER
1240	1flr	L	28	248	1.6e-65	-0.10	0.60		4-4-20 (IG*G2A=KAPPA=) FAB FRAGMENT, 1FLR 5 CHAIN: L, H, 1FLR 6	IMMUNOGLOBULIN
1240	1fv1	B	47	245	9.6e-47	-0.24	0.04		MAJOR HISTOCOMPATIBILITY COMPLEX ALPHA CHAIN; CHAIN: A, D; MAJOR HISTOCOMPATIBILITY COMPLEX BETA CHAIN; CHAIN: B, E; MYELIN BASIC PROTEIN; CHAIN: C, F;	IMMUNE SYSTEM MHC CLASS II DR2A
1240	1fvd	A	27	248	9.6e-63			84.25	IMMUNOGLOBULIN FAB FRAGMENT OF HUMANIZED ANTIBODY 4D5, VERSION 4 1FVD 3	
1240	1ghf	L	27	244	3.2e-59			85.44	ANTI-ANTI-IDIO TYPE GH1002 FAB FRAGMENT; CHAIN: L, H	ANTIBODY FAB FRAGMENT
1240	1gpo	L	25	244	1.6e-63			85.20	ANTIBODY M41; CHAIN: L, H, M, I;	IMMUNOGLOBULIN PROTEIN ENGINEERING, ANTIBODY DESIGN, IMMUNOGLOBULIN 2 STRUCTURE, ANTIGEN-BINDING SITE, CANONICAL CONFORMATION, 3 COMPLEMENTARITY-DETERMINING REGION

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1240	1hil	A	25	244	4.8e-65			83.07	IMMUNOGLOBULIN IGG2A FAB FRAGMENT (FAB 17/9) 1HIL 3	
1240	1iao	B	74	244	6.4e-46	-0.21	0.13		MHC CLASS II I-AD; CHAIN: A, B;	MHC II MHC II, CLASS II MHC, I-A, OVALBUMIN PEPTIDE
1240	1ieb	B	74	243	1.3e-44	-0.36	0.00		MHC CLASS II I-EK; CHAIN: A, B, C, D;	HISTOCOMPATIBILITY ANTIGEN HISTOCOMPATIBILITY ANTIGEN
1240	1ifh	L	25	244	4.8e-65			83.18	IMMUNOGLOBULIN IGG2A FAB FRAGMENT (FAB 17/9) COMPLEX WITH PEPTIDE OF 1IFH 3 INFLUENZA HEMAGGLUTININ HA1 (STRAIN X47) (RESIDUES 101-107) 1IFH 4	
1240	1incp	L	28	248	6.4e-66	-0.19	0.21		IMMUNOGLOBULIN IMMUNOGLOBULIN FAB FRAGMENT (MCP/PC\$603) 1MCP 4	
1240	1imfb	L	27	234	3.2e-50	0.01	0.55		IMMUNOGLOBULIN FAB FRAGMENT (MURINE SE155-4) COMPLEX WITH HEPTASACCHARIDE 1MFB 3 B: GAL(1-2)MAN(1-4)RAM(1-3)GAL(1-2)[ABE(1-3)]MAN(1-4)RAM 1MFB 4	
1240	1inca	L	28	248	1.4e-65	-0.10	0.52		HYDROLASE(O-GLYCOSYL)	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									N9 NEURAMINIDASE-NC41 (E.C.3.2.1.18) COMPLEX WITH FAB INCA 3	
1240	1neu		42	140	1.9e-05	0.21	0.09		MYELIN P0 PROTEIN; CHAIN: NULL;	STRUCTURAL PROTEIN MYELIN, STRUCTURAL PROTEIN, GLYCOPROTEIN, TRANSMEMBRANE, PHOSPHORYLATION, IMMUNOGLOBULIN FOLD, SIGNAL, MYELIN 2 MEMBRANE ADHESION MOLECULE
1240	1nsh	L	28	247	6.4e-66	-0.07	0.25		IGG FAB (IGG1, KAPPA); INSN 4 CHAIN: L, H; INSN 5 STAPHYLOCOCCAL NUCLEASE; INSN 9 CHAIN: S; INSN 10	COMPLEX IMMUNOGLOBULIN/HYDROLASE; N10 FAB IMMUNOGLOBULIN; INSN 7 STAPHYLOCOCCAL NUCLEASE RIBONUCLEASE, INSN 11 IMMUNOGLOBULIN, STAPHYLOCOCCAL NUCLEASE INSN 25
1240	1qlr	A	28	248	1.6e-65	-0.20	0.16		IGM KAPPA CHAIN V-III (KAU COLD AGGLUTININ); CHAIN: A, C; IGM FAB REGION IV- J(H4)-C (KAU COLD AGGLUTININ); CHAIN: B, D;	IMMUNOGLOBULIN IMMUNOGLOBULIN, AUTOANTIBODY, COLD AGGLUTININ, HUMAN IGM 2 FAB FRAGMENT
1240	1i24	A	25	240	6.4e-59			83.15	IGG3-KAPPA ANTIBODY (LIGHT CHAIN); CHAIN: A, C; IGG3-KAPPA ANTIBODY	IMMUNE SYSTEM PRELIMINARY, IMMUNE SYSTEM

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1240	1sbs	L	25	248	1.1e-66			83.76	(HEAVY CHAIN); CHAIN: B, D; MONOCLONAL ANTIBODY 3A2; CHAIN: H, L;	MONOCLONAL ANTIBODY MONOCLONAL ANTIBODY, FAB-FRAGMENT, REPRODUCTION
1240	1sbs	L	28	248	1.1e-66	-0.28	0.30		MONOCLONAL ANTIBODY 3A2; CHAIN: H, L;	MONOCLONAL ANTIBODY MONOCLONAL ANTIBODY, FAB-FRAGMENT, REPRODUCTION
1240	1sm3	L	28	234	8e-51	0.24	0.96		SM3 ANTIBODY; CHAIN: L, H; PEPTIDE EPTOPE; CHAIN: P;	COMPLEX (ANTIBODY/PEPTIDE EPTOPE) ANTIBODY, PEPTIDE ANTIGEN, ANTITUMOR ANTIBODY, 2 COMPLEX (ANTIBODY/PEPTIDE EPTOPE)
1240	1vge	L	27	248	4.8e-63			83.26	TR1.9 FAB; CHAIN: L, H;	IMMUNOGLOBULIN TR1.9, ANTI-THYROID PEROXIDASE, AUTOANTIBODY, 2 IMMUNOGLOBULIN
1240	2fgw	L	27	248	3.2e-63			83.51	IMMUNOGLOBULIN FAB FRAGMENT OF A HUMANIZED VERSION OF THE ANTI-CD18 2FGW 3 ANTIBODY Y52 (HUH52-OZ FAB) 2FGW 4	
1240	2gfb	A	27	248	9.6e-62			81.62	IMMUNOGLOBULIN IGG2A FAB FRAGMENT (CNJ206) 2GFB 3	
1240	2hmi	C	27	248	8e-61			84.84	HIV-1 REVERSE	COMPLEX (RT/DNA/FAB) HIV-1

264.

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									TRANSCRIPTASE; CHAIN: A, B; MONOCLONAL ANTIBODY 28; CHAIN: C, D; DNA; CHAIN: E, F;	RT, FAB 28; AIDS, HIV-1, RT, POLYMERASE
1240	32c2	A	28	248	6.4e-66	-0.19	0.42		IGG1 ANTIBODY 32C2; CHAIN: A; IGG1 ANTIBODY 32C2; CHAIN: B;	IMMUNE SYSTEM FAB, ANTIBODY, AROMATASE, P450
1240	6fab	L	25	248	3.2e-62			82.05	IMMUNOGLOBULIN ANTIGEN-BINDING FRAGMENT OF THE MURINE ANTI-PHENYLARSONATE 6FAB 3 ANTIBODY 36-71, "FAB 36-71" 6FAB 4	
1240	7fab	L	30	236	3.2e-57	-0.01	0.45		IMMUNOGLOBULIN IMMUNOGLOBULIN FAB' NEW (LAMBDA LIGHT CHAIN) 7FAB 3	
1240	8fab	A	29	236	1.1e-58	0.03	0.52		IMMUNOGLOBULIN FAB FRAGMENT FROM HUMAN IMMUNOGLOBULIN IGG1 (LAMBDA, HIL) 8FAB 3	
1247	1cok	A	6	67	0.0048	-0.62	0.04		SECOND SPLICE VARIANT P73; CHAIN: A;	GENE REGULATION P73 SAM-LIKE DOMAIN, GENE REGULATION
1258	1htr	P	18	59	0.00011	-0.71	0.65		ASPARTYL PROTEASE PROGASTRICIN	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									(PEPSINOGEN C) (E.C.3.4.23.3) 1HTR 3 1HTR 87	
1258	1ht	P	18	60	2.9e-13	-0.87	0.99		ASPARTYL PROTEASE PROGASTRICIN (PEPSINOGEN C) (E.C.3.4.23.3) 1HTR 3 1HTR 87	
1258	1qrp	E	61	239	2.9e-35	-0.34	0.19		PEPSIN 3A; CHAIN: E; IVA-VAL-VAL-LEU(P)-(O)PHE-ALA-ALA-OME; CHAIN: I; 1HTR 3 1HTR 87	HYDROLASE/HYDROLASE INHIBITOR ASPARTIC PROTEINASE, PHOSPHONATE INHIBITOR, TRANSITION 2 STATE ANALOGUE
1258	3cms		59	239	4.8e-34	-0.30	0.99		HYDROLASE (ACID PROTEINASE) CHYMOSIN B (FORMERLY KNOWN AS RENNIN) (E.C.3.4.23.4) MUTANT 3CMS 3 WITH VAL 111 REPLACED BY PHE (V111F) 3CMS 4	
1258	3psg		19	239	1.3e-39	-0.29	0.05		HYDROLASE (ACID PROTEINASE ZYMOGEN) PEPINOGEN 3PSG 3	
1258	3psg		19	239	1.9e-45	-0.10	0.95		HYDROLASE (ACID PROTEINASE ZYMOGEN) PEPINOGEN 3PSG 3	
1258	4pep		61	239	4.3e-34	-0.30	0.37		HYDROLASE (ACID PROTEINASE) PEPIN (E.C.3.4.23.1) 4PEP 4	



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1267	1am5		70	190	0.0019	-0.56	0.30		PEPSIN; CHAIN: NULL;	ASPARTYL PROTEASE ACID PROTEINASE; ASPARTYL PROTEASE, ACID PROTEINASE, HYDROLASE
1276	1ebo	A	96	134	2.9e-07	-0.78	0.07		EBOLA VIRUS ENVELOPE PROTEIN CHIMERA CONSISTING CHAIN: A, B, C, D, E, F;	VIRUS/VIRAL PROTEIN MEMBRANE FUSION SUBUNIT, VIRUS/VIRAL PROTEIN
1288	1b8q	A	18	73	1.7e-06	-0.34	0.12		NEURONAL NITRIC OXIDE SYNTHASE; CHAIN: A; HEPTAPEPTIDE; CHAIN: B;	OXIDOREDUCTASE PDZ DOMAIN, NNOS, NITRIC OXIDE SYNTHASE
1288	1be9	A	33	72	0.0013	0.27	0.12		PSD-95; CHAIN: A; CRIFT; CHAIN: B;	PEPTIDE RECOGNITION, PEPTIDE RECOGNITION, PROTEIN LOCALIZATION
1288	1kwa	A	20	73	1.7e-07	-0.30	0.48		HCASK/LIN-2 PROTEIN; CHAIN: A, B;	KINASE HCASK, GLGF REPEAT, DHR, PDZ DOMAIN, NEUREXIN, SYNDECAN, RECEPTOR CLUSTERING, KINASE
1288	1pdr		26	73	0.00017	0.13	0.39		HUMAN DISCS LARGE PROTEIN; CHAIN: NULL;	SIGNAL TRANSDUCTION HDLG, DHR3 DOMAIN, SIGNAL TRANSDUCTION, SH3 DOMAIN, REPEAT
1288	1gau	A	20	73	4.3e-06	-0.07	0.27		NEURONAL NITRIC OXIDE SYNTHASE (RESIDUES 1-130); CHAIN: A;	OXIDOREDUCTASE BETA-FINGER
1288	1gav	A	26	73	8.6e-07	0.25	0.83		ALPHA-1 SYNTROPHIN	MEMBRANE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1288	1qlc	A	32	75	3.4e-07	-0.01	0.13		(RESIDUES 77-171); CHAIN: A; NEURONAL NITRIC OXIDE SYNTHASE (RESIDUES 1-130); CHAIN: B;	PROTEIN/OXIDOREDUCTASE BETA-FINGER, HETERODIMER
1288	3pdz	A	20	75	4.3e-06	0.27	0.78		POSTSYNAPTIC DENSITY PROTEIN 95; CHAIN: A; TYROSINE PHOSPHATASE (PTP-BAS, TYPE 1); CHAIN: A;	PEPTIDE RECOGNITION PSD-95; PDZ DOMAIN, NEURONAL NITRIC OXIDE SYNTHASE, NMDA RECEPTOR 2 BINDING HYDROLASE PDZ DOMAIN, HUMAN PHOSPHATASE, HPTP1E, PTP-BAS, SPECIFICITY 2 OF BINDING
1302	lbyr	A	75	170	1.7e-05	-0.14	0.25		ENDONUCLEASE; CHAIN: A;	ENDONUCLEASE ENDONUCLEASE, PHOSPHODIESTERASE,
1312	lrl		19	70	0.0003	-0.26	0.00		HYDROLASE/ENDORIBONUCLEASE) RIBONUCLEASE H (E.C.3.1.26.4) IRL 3	
1313	ljnk		1	52	0.0086	-0.89	0.19		C-JUN N-TERMINAL KINASE; CHAIN: NULL;	TRANSFERASE JNK3; TRANSFERASE, JNK3 MAP KINASE, SERINE/THREONINE PROTEIN 2 KINASE
1320	lcp2	A	25	73	0.0021	-0.67	0.06		NITROGENASE IRON PROTEIN; CHAIN: A, B;	OXIDOREDUCTASE CP2; OXIDOREDUCTASE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1320	2nlp	A	27	65	0.00048	-0.47	0.52		NITROGENASE IRON PROTEIN; CHAIN: A, B;	NITROGENASE IRON PROTEIN HEADER CONNECT LINK
1320	2nlp	B	27	65	0.00048	-0.37	0.46		NITROGENASE IRON PROTEIN; CHAIN: A, B;	IRON PROTEIN IRON PROTEIN, OXIDOREDUCTASE
1341	1bpv		212	314	8.6e-16	0.19	-0.08		TTIN; CHAIN: NULL;	CONNECTIN A71, CONNECTIN; TTIN, CONNECTIN, FIBRONECTIN TYPE III
1341	1bqu	A	1	221	2.6e-16			50.84	GP130; CHAIN: A, B;	SIGNALING PROTEIN CYTOKINE RECEPTOR, GLYCOPROTEIN 130, GP130, INTERLEUKINE 6 2 RECEPTOR BETA SUBUNIT, SIGNALING PROTEIN
1341	1cfb		104	313	1.7e-36			84.74	NEURAL ADHESION MOLECULE DROSOPHILA NEUROGLIAN (CHYMOTRYPTIC FRAGMENT CONTAINING THE 1CFB 3 TWO AMINO PROXIMAL FIBRONECTIN TYPE III REPEATS 1CFB 4 (RESIDUES 610 - 814)) 1CFB 5	
1341	1cfb		106	312	1.7e-36	-0.02	0.12		NEURAL ADHESION MOLECULE DROSOPHILA NEUROGLIAN	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
									(CHYMOTRYPTIC FRAGMENT CONTAINING THE 1CFB 3 TWO AMINO PROXIMAL FIBRONECTIN TYPE III REPEATS 1CFB 4 (RESIDUES 610 - 814)) 1CFB 5	
1341	1cfb		50	207	2.6e-26	-0.26	0.11		NEURAL ADHESION MOLECULE DROSOPHILA NEUROGLIAN (CHYMOTRYPTIC FRAGMENT CONTAINING THE 1CFB 3 TWO AMINO PROXIMAL FIBRONECTIN TYPE III REPEATS 1CFB 4 (RESIDUES 610 - 814)) 1CFB 5	
1341	1cfb		6	205	1.1e-23	-0.26	0.01		NEURAL ADHESION MOLECULE DROSOPHILA NEUROGLIAN (CHYMOTRYPTIC FRAGMENT CONTAINING THE 1CFB 3 TWO AMINO PROXIMAL FIBRONECTIN TYPE III REPEATS 1CFB 4 (RESIDUES 610 - 814)) 1CFB 5	
1341	166f	B	111	309	3.2e-20	0.13	-0.01		PLACENTAL LACTOGEN; CHAIN: A; PROLACTIN RECEPTOR; CHAIN: B, C;	HORMONE/GROWTH FACTOR/HORMONE RECEPTOR 4-HELICAL BUNDLE, ALPHA HELICAL BUNDLE, TERNARY

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										COMPLEX, FN 2 III DOMAINS, BETA SHEET DOMAINS, CYTOKINE-RECEPTOR COMPLEX
1341	1mf		13	354	3.2e-33			76.52	FIBRONECTIN; 1FNF 6 CHAIN: NULL; 1FNF 7	CELL ADHESION PROTEIN RGD, EXTRACELLULAR MATRIX
1341	1fhh	A	112	340	8.6e-26	-0.02	0.27		FIBRONECTIN; CHAIN: A;	HEPARIN AND INTEGRIN BINDING HEPARIN AND INTEGRIN BINDING
1341	1fhh	A	12	300	1.6e-26	-0.29	0.40		FIBRONECTIN; CHAIN: A;	HEPARIN AND INTEGRIN BINDING HEPARIN AND INTEGRIN BINDING
1341	1fhh	A	15	310	1.6e-26			81.90	FIBRONECTIN; CHAIN: A;	HEPARIN AND INTEGRIN BINDING HEPARIN AND INTEGRIN BINDING
1341	1mfh		11	200	3.2e-17	-0.08	0.30		FIBRONECTIN; CHAIN: NULL;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN, RGD, EXTRACELLULAR MATRIX, 2
										HEPARIN-BINDING, GLYCOPROTEIN
1341	1mfh		110	310	4.3e-27			54.17	FIBRONECTIN; CHAIN: NULL;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN, RGD, EXTRACELLULAR MATRIX, 2
										HEPARIN-BINDING, GLYCOPROTEIN
1341	1mfh		112	310	4.3e-27	0.09	0.49		FIBRONECTIN; CHAIN: NULL;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN, RGD,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										EXTRACELLULAR MATRIX, 2 HEPARIN-BINDING, GLYCOPROTEIN
1341	1mfn		214	351	6.4e-18	-0.37	0.03		FIBRONECTIN; CHAIN: NULL;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN, RGD, EXTRACELLULAR MATRIX, 2 HEPARIN-BINDING, GLYCOPROTEIN
1341	1mfn		49	207	4.3e-18	0.01	0.18		FIBRONECTIN; CHAIN: NULL;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN, RGD, EXTRACELLULAR MATRIX, 2 HEPARIN-BINDING, GLYCOPROTEIN
1341	1qg3	A	11	158	1.3e-16	0.08	-0.17		INTEGRIN BETA-4 SUBUNIT; CHAIN: A, B;	STRUCTURAL PROTEIN INTEGRIN, HEMIDESMOSOME, FIBRONECTIN, CARCINOMA, STRUCTURAL 2 PROTEIN
1341	1qg3	A	110	312	8.6e-31			96.78	INTEGRIN BETA-4 SUBUNIT; CHAIN: A, B;	STRUCTURAL PROTEIN INTEGRIN, HEMIDESMOSOME, FIBRONECTIN, CARCINOMA, STRUCTURAL 2 PROTEIN
1341	1qg3	A	111	304	3.2e-17	0.03	0.71		INTEGRIN BETA-4 SUBUNIT; CHAIN: A, B;	STRUCTURAL PROTEIN INTEGRIN, HEMIDESMOSOME, FIBRONECTIN, CARCINOMA, STRUCTURAL 2 PROTEIN
1341	1qg3	A	112	310	8.6e-31	0.13	0.96		INTEGRIN BETA-4 SUBUNIT; CHAIN: A, B;	STRUCTURAL PROTEIN INTEGRIN, HEMIDESMOSOME, FIBRONECTIN, CARCINOMA,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Pst Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1341	1qg3	A	214	351	8e-24	-0.11	0.48		INTEGRIN BETA-4 SUBUNIT; CHAIN: A, B;	STRUCTURAL 2 PROTEIN
1341	1qg3	A	50	209	3e-22	-0.11	0.45		INTEGRIN BETA-4 SUBUNIT; CHAIN: A, B;	STRUCTURAL PROTEIN INTEGRIN, HEMIDESMOSOME, FIBRONECTIN, CARCINOMA, STRUCTURAL 2 PROTEIN
1341	1qr4	A	111	307	4.3e-28			56.55	TENASCIN; CHAIN: A, B;	STRUCTURAL PROTEIN TENASCIN, FIBRONECTIN TYPE-III, HEPARIN, EXTRACELLULAR 2 MATRIX, ADHESION, FUSION PROTEIN, STRUCTURAL PROTEIN
1341	1qr4	A	112	305	4.3e-28	0.24	0.48		TENASCIN; CHAIN: A, B;	STRUCTURAL PROTEIN TENASCIN, FIBRONECTIN TYPE-III, HEPARIN, EXTRACELLULAR 2 MATRIX, ADHESION, FUSION PROTEIN, STRUCTURAL PROTEIN
1341	2fmb	A	209	303	4.3e-15	0.06	0.30		FIBRONECTIN; CHAIN: A;	PROTEIN BINDING ED-B, FIBRONECTIN, TYPEIII DOMAIN, ANGIOGENESIS, PROTEIN 2 BINDING
1363	1c3p	A	91	159	4.3e-17	-0.34	0.06		HDLP (HISTONE DEACETYLASE-LIKE	LYASE ALPHA/BETA FOLD, LYASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1363	1c3p	A	91	162	3.2e-15	0.28	0.27		PROTEIN; CHAIN: A; HDLP (HISTONE DEACETYLASE-LIKE PROTEIN); CHAIN: A;	LYASE ALPHA/BETA FOLD, LYASE
1364	1d1d	A	117	163	1.1e-09	-0.19	0.90		CAPSID PROTEIN; CHAIN: A;	VIRUS/VIRAL PROTEIN TWO INDEPENDENT DOMAINS HELICAL BUNDLES, VIRUS/VIRAL PROTEIN
1364	1em9	A	107	163	6.4e-11	-0.14	0.48		GAG POLYPROTEIN CAPSID PROTEIN P27; CHAIN: A, B;	VIRUS/VIRAL PROTEIN
1364	1em9	B	115	163	3.2e-10	-0.17	0.82		GAG POLYPROTEIN CAPSID PROTEIN P27; CHAIN: A, B;	VIRUS/VIRAL PROTEIN
1364	1qj	B	122	163	8.6e-06	-0.13	0.90		HIS TAG; CHAIN: A; HTLV-I CAPSID PROTEIN; CHAIN: B;	VIRUS/VIRAL PROTEIN HTLV-I, CAPSID PROTEIN, RETROVIRUS, TWO-DOMAIN PROTEIN, 2 ALPHA HELICAL PROTEIN, HETERONUCLEAR NMR SPECTROSCOPY, 3 VIRUS/VIRAL PROTEIN
1379	1eze	A	20	54	2.1e-06	-0.82	0.18		CHOLESTERYL ESTER TRANSFERASE INHIBITOR PROTEIN; CHAIN: A;	TRANSFERASE INHIBITOR CETIP, APOLIPROTEIN C-I, APO-CI; AMPHIPATHIC HELIX
1379	1ioj		18	74	3.2e-21			61.29	APOC-I; CHAIN: NULL;	APOLIPROTEIN, AMPHIPATHIC HELIX, LIPID ASSOCIATION, LCAT 2



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1379	1ioj		26	74	3.2e-21	-0.55	0.54		APOC-I; CHAIN: NULL;	ACTIVATION APOLIPROTEIN APOLIPROTEIN, AMPHIPATHIC HELIX, LIPID ASSOCIATION, LCAT 2 ACTIVATION
1379	1opp		20	55	4.3e-07	-0.64	0.12		APOLIPROTEIN C-I; CHAIN: NULL;	APOLIPROTEIN APO-CI; APOLIPROTEIN, AMPHIPATHIC HELIX, LIPID ASSOCIATION, LCAT 2 ACTIVATION
1396	1dan	L	259	333	9.6e-10	0.15	0.11		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLOROMETHYLKETONE (DEFCMK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO- FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
1396	1dva	L	259	333	9.6e-10	0.20	0.10		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE COMPLEX
1396	1fvl		245	313	3.2e-14	0.38	0.95		FLAVORIDIN; 1FVL 4 CHAIN: NULL 1FVL 5	BLOOD COAGULATION INHIBITOR GP IIB/IIIA ANTAGONIST 1FVL 9
1396	1fvl		245	316	8.6e-24			69.59	FLAVORIDIN; 1FVL 4 CHAIN:	BLOOD COAGULATION

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									NULL 1FVL 5	INHIBITOR GP IIB/IIIA ANTAGONIST 1FVL 9
1396	1fvl		246	318	8.6e-24	0.30	0.39		FLAVORDIN; 1FVL 4 CHAIN: NULL 1FVL 5	BLOOD COAGULATION INHIBITOR GP IIB/IIIA ANTAGONIST 1FVL 9
1396	1klo		218	360	1.3e-11	0.13	-0.14		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
1396	1kst		245	313	4.8e-15	0.17	0.21		AGGREGATION INHIBITOR, GP ANTAGONIST KISTRIN (NMR, 8 STRUCTURES) IKST 3	
1396	1kst		245	314	4.3e-23			68.76	AGGREGATION INHIBITOR, GP ANTAGONIST KISTRIN (NMR, 8 STRUCTURES) IKST 3	
1396	1kst		246	313	4.3e-23	0.54	0.58		AGGREGATION INHIBITOR, GP ANTAGONIST KISTRIN (NMR, 8 STRUCTURES) IKST 3	
1396	1kta	L	261	333	9.6e-10	0.30	-0.15		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
1396	2ech		273	322	3.4e-18			50.38	BLOOD COAGULATION INHIBITOR ECHISTATIN (NMR, 8 STRUCTURES) 2ECH 3	
1396	2ech		274	322	3.4e-18	0.34	0.28		BLOOD COAGULATION INHIBITOR ECHISTATIN (NMR, 8 STRUCTURES) 2ECH 3	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1406	1dgm	A	10	66	0.0059	-0.64	0.01		ADENOSINE KINASE; CHAIN: A;	TRANSFERASE TOXOPLASMA GONDI, ADENOSINE KINASE, PURINE METABOLISM
1409	1bfa	A	45	71	0.0086	-0.11	0.17		SSO7D; CHAIN: A; DNA; CHAIN: B, C;	COMPLEX (DNA-BINDING PROTEIN/DNA) DNA BINDING PROTEIN, HYPERTHERMOPHILE, ACHAEBACTERIA, 2 COMPLEX (DNA-BINDING PROTEIN/DNA)
1413	1e2q	A	26	108	8e-05	-0.18	0.51		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN: A;	IMMUNE SYSTEM FC-EPSILON R1-ALPHA; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IGE-BINDING 2 PROTEIN
1413	1fcg	A	11	108	8.6e-06	-0.08	0.37		FC RECEPTOR FC(GAMMA)RIIA; CHAIN: A;	IMMUNE SYSTEM, MEMBRANE PROTEIN CD32; FC RECEPTOR, IMMUNOGLOBULIN, LEUKOCYTE, CD32
1413	1fhl	A	33	111	8.6e-07	-0.12	0.13		LOW AFFINITY IMMUNOGLOBULIN GAMMA FC REGION CHAIN: A;	IMMUNE SYSTEM RECEPTOR BETA SANDWICH, IMMUNOGLOBULIN-LIKE, RECEPTOR
1413	1nkr		22	109	3.2e-29	-0.16	0.68		P58-CL42 KIR; CHAIN: NULL;	INHIBITORY RECEPTOR KILLER CELL INHIBITORY RECEPTOR;

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1413	1nkt		4	109	4.3e-12	-0.13	0.47		P58-CL42 KIR; CHAIN: NULL;	INHIBITORY RECEPTOR, NATURAL KILLER CELLS, IMMUNOLOGICAL 2 RECEPTORS, IMMUNOGLOBULIN FOLD
1413	2dli	A	22	108	1.1e-27	-0.05	0.17		MHC CLASS I NK CELL RECEPTOR PRECURSOR; CHAIN: A;	IMMUNOGLOBULIN FOLD IMMUNE SYSTEM P58 NATURAL KILLER CELL RECEPTOR; KIR, NATURAL KILLER RECEPTOR, INHIBITORY RECEPTOR, 2 IMMUNOGLOBULIN
1413	2dli	A	6	108	1.3e-07	-0.11	0.27		MHC CLASS I NK CELL RECEPTOR PRECURSOR; CHAIN: A;	IMMUNE SYSTEM P58 NATURAL KILLER CELL RECEPTOR; KIR, NATURAL KILLER RECEPTOR, INHIBITORY RECEPTOR, 2 IMMUNOGLOBULIN
1415	1gdn	A	16	107	0.0038	0.20	0.27		OXIDOREDUCTASE(CHOH (D)-NAD(P) <sup>+</sup> (A)) D-GLYCERATE	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
									DEHYDROGENASE (APO FORM) (E.C.1.1.29) 1GDH 3	
1416	1dun		109	230	4.8e-24			54.22	DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE; CHAIN: NULL;	HYDROLASE DUTPASE, DUTP PYROPHOSPHATASE; HYDROLASE, DUTPASE, EIAY, TRIMERIC ENZYME, ASPARTYL PROTEASE
1416	1dun		123	219	4.8e-24	0.16	0.98		DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE; CHAIN: NULL;	HYDROLASE DUTPASE, DUTP PYROPHOSPHATASE; HYDROLASE, DUTPASE, EIAY, TRIMERIC ENZYME, ASPARTYL PROTEASE
1416	1euw	A	102	222	9.6e-26	0.28	0.28		DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE; CHAIN: A;	HYDROLASE DUTPASE; JELLY ROLL, MERCURY DERIVATIVE
1416	1f7d	A	107	214	4.8e-27	0.25	0.88		POLY(POLYPROTEIN); CHAIN: A, B;	VIRUS/VIRAL PROTEIN EIGHT STRANDED BETA-BARREL
1416	1f7r	A	107	235	1.1e-31	0.28	0.29		POLY(POLYPROTEIN); CHAIN: A;	VIRUS/VIRAL PROTEIN EIGHT STRANDED BETA BARREL PROTEIN
1426	1a12	A	16	119	1.3e-12	0.51	0.39		REGULATOR OF CHROMOSOME CONDENSATION 1; CHAIN: A, B, C;	GUANINE NUCLEOTIDE EXCHANGE FACTOR RCC1; GUANINE NUCLEOTIDE EXCHANGE FACTOR, GEF, RAN, 2 RAS-LIKE NUCLEAR GTP

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1426	1a12	A	32	121	9.6e-19	0.32	0.39		REGULATOR OF CHROMOSOME CONDENSATION 1; CHAIN: A, B, C;	BINDING PROTEIN HEADER TER GUANINE NUCLEOTIDE EXCHANGE FACTOR RCC1; GUANINE NUCLEOTIDE EXCHANGE FACTOR, GEF, RAN, 2 RAS-LIKE NUCLEAR GTP BINDING PROTEIN HEADER TER
1434	1du8	A	195	318	3.2e-38	0.02	0.78		SUREACTANT PROTEIN A; CHAIN: A;	MEMBRANE PROTEIN SP-A; SP-A:PHOSPHOLIPID MOLECULAR COMPLEX
1435	2fc6	A	27	71	3.4e-06	-0.89	0.09		FC GAMMA RIIB; CHAIN: A;	IMMUNE SYSTEM CD32; RECEPTOR, FC, CD32, IMMUNE SYSTEM
1440	1ap0		89	155	4.8e-21	-0.14	0.01		MODIFIER PROTEIN 1; CHAIN: NULL;	CHROMATIN-BINDING MOMOD1, HETEROCHROMATIN PROTEIN 1; CHROMATIN-BINDING, PROTEIN INTERACTION MOTIF, ALPHA+BETA
1449	1fa2	A	15	101	1.1e-29	-0.28	0.00		PEPTIDE METHIONINE SULFOXIDE REDUCTASE; CHAIN: A, B;	OXIDOREDUCTASE OXIDOREDUCTASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1456	1qoj	B	57	104	0.0017	0.07	0.35		UVRB; CHAIN: A, B;	DNA EXCISION REPAIR NUCLEOTIDE EXCISION REPAIR, X-RAY CRYSTALLOGRAPHY, UVRB 2 PROTEIN, UVRB-C INTERACTION
1458	1bkt		183	210	3.2e-11	-0.68	0.90		FK506 BINDING PROTEIN; CHAIN: NULL;	ISOMERASE FKBP; ISOMERASE, ROTAMASE
1458	1c9h	A	183	210	4.8e-11	-0.80	0.78		FKBP12.6; CHAIN: A;	IMMUNE SYSTEM CALCINEURIN, FKBP12, RAPAMYCIN, COMPLEX, RYANODINE RECEPTOR
1463	1f5a	A	88	264	4.8e-35	-0.41	0.13		POLY(A) POLYMERASE; CHAIN: A;	TRANSFERASE MRNA PROCESSING, TRANSFERASE, TRANSCRIPTION, RNA- BINDING, 2
1463	1fa0	A	85	264	1.3e-28	-0.31	0.24		POLY(A)-POLYMERASE; CHAIN: A, B;	PHOSPHORYLATION, NUCLEAR PROTEIN, ALTERNATIVE SPLICING 3 HELICAL TURN MOTIF, NUCLEOTIDYL TRANSFERASE CATALYTIC DOMAIN
1463	1kay	A	119	164	0.0078	-0.70	0.17		KANAMYCIN	TRANSFERASE POLYMERASE, NUCLEOTIDYL TRANSFERASE TRANSFERASE KINASE;

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									NUCLEOTIDYLTRANSFERASE ; CHAIN: A, B;	ANTIBIOTIC RESISTANCE, TRANSFERASE, PLASMID
1467	1faq		130	182	4.8e-13	-0.05	0.04		RAF-1; CHAIN: NULL;	SERINE/THREONINE PROTEIN KINASE TRANSFERASE, SERINE/THREONINE-PROTEIN KINASE, 2 PROTO-ONCOGENE, ZINC, ATP-BINDING, PHORBOL-ESTER BINDING
1467	1ptq		130	179	1.1e-17	0.08	0.23		PROTEIN KINASE C DELTA TYPE; IPTQ 4	PHOSPHOTRANSFERASE
1468	1av1	A	23	228	1.3e-09			57.74	APOLIPOPROTEIN A-I; CHAIN: A, B, C, D;	LIPID TRANSPORT APO A-I; LIPOPROTEIN, LIPID TRANSPORT, CHOLESTEROL METABOLISM, 2
1468	1cun	A	10	220	2.1e-13			52.54	ALPHA SPECTRIN; CHAIN: A, B, C;	ATHEROSCLEROSIS, HDL, LCAT-ACTIVATION
1468	1dn1	B	99	236	1.3e-08	-0.00	-0.19		SYNTAXIN BINDING PROTEIN 1; CHAIN: A; SYNTAXIN 1A; CHAIN: B;	STRUCTURAL PROTEIN TWO REPEATS OF SPECTRIN, ALPHA HELICAL LINKER REGION, 2 2 TANDEM 3-HELIX COILED-COILS, STRUCTURAL PROTEIN
1468	1elr	A	82	228	0.00086	0.26	0.22		TPR2A-DOMAIN OF HOP; CHAIN: A; HSP90-PEPTIDE	ENDOCTOSIS/EXOCYTOSIS NSEC; PROTEIN-PROTEIN COMPLEX, MULTI-SUBUNIT CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX,



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1468	1ez3	A	107	229	8.6e-10	0.18	-0.19		MEEVD; CHAIN: B; SYNTAXIN-1A; CHAIN: A, B, C;	HELICAL REPEAT, HSP90, 2 PROTEIN BINDING ENDOCYTOSIS/EXOCYTOSIS SYNAPTOTAGMIN ASSOCIATED 35 KDA PROTEIN, P35A, THREE HELIX BUNDLE
1468	1qsa	A	34	237	3e-20	0.08	-0.19		SOLUBLE LYTIC TRANSGLYCOSYLASE SLT70; CHAIN: A;	TRANSFERASE ALPHA-SUPERHELIX, TRANSFERASE
1468	1quu	A	33	275	8.6e-20			56.44	HUMAN SKELETAL MUSCLE ALPHA-ACTININ 2; CHAIN: A;	CONTRACTILE PROTEIN TRIPLE-HELIX COILED COIL, CONTRACTILE PROTEIN
1468	1quu	A	34	234	8.6e-20	0.03	-0.14		HUMAN SKELETAL MUSCLE ALPHA-ACTININ 2; CHAIN: A;	CONTRACTILE PROTEIN TRIPLE-HELIX COILED COIL, CONTRACTILE PROTEIN
1474	1a06		7	167	4.8e-33	-0.52	0.76		CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE; CHAIN: NULL;	KINASE KINASE, SIGNAL TRANSDUCTION, CALCIUM/CALMODULIN
1474	1apm	E	9	185	3.2e-51	0.03	0.87		TRANSFERASE(PHOSPHOTRANSFERASE) \$C-/AMP\$-DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (\$C/APK\$) IAPM3 (CATALYTIC SUBUNIT) "ALPHA" ISOENZYME MUTANT WITH SER 139 IAPM4 REPLACED BY ALA	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1474	1cmk	E	9	185	3.2e-52	-0.05	0.89		(S139A\$) COMPLEX WITH THE PEPTIDE IAPM 5 INHIBITOR PKI(5-24) AND THE DETERGENT MBGA-8 IAPM 6	
1474	1cnp	E	9	185	3.2e-52	0.04	0.80		PHOSPHOTRANSFERASE CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT 1CMK 3 (E.C.2.7.1.37) 1CMK 4 TRANSFERASE(PHOSPHOTRANSFERASE) CAMP-DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (CAPK) 1CTP 3 (CATALYTIC SUBUNIT) 1CTP 4	
1474	1E3m	C	7	181	1.4e-46	-0.28	0.09		SERINE/THREONINE-PROTEIN KINASE PAK-ALPHA: CHAIN: A, B; SERINE/THREONINE-PROTEIN KINASE PAK-ALPHA: CHAIN: C, D; TWITCHIN; CHAIN: NULL;	TRANSFERASE KINASE DOMAIN, AUTOINHIBITORY FRAGMENT, HOMODIMER
1474	1koa		10	160	4.8e-31	0.02	0.57		TWITCHIN; CHAIN: NULL;	KINASE KINASE, TWITCHIN, INTRASTERIC REGULATION
1474	1kob	A	10	162	3.2e-31	0.05	0.18		TWITCHIN; CHAIN: A, B;	KINASE KINASE, TWITCHIN, INTRASTERIC REGULATION
1474	1phk		7	161	3.2e-43	-0.11	0.39		PHOSPHORYLASE KINASE; CHAIN: NULL;	KINASE RABBIT MUSCLE PHOSPHORYLASE KINASE; GLYCOGEN METABOLISM, TRANSFERASE, SERINE/THREONINE-PROTEIN, 2

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1474	1pme		22	160	6.4e-31	-0.12	0.36		ERK2; CHAIN: NULL;	KINASE, ATP-BINDING, CALMODULIN-BINDING
1474	1qpc	A	10	179	4.8e-29	0.19	0.55		LCK KINASE; CHAIN: A;	SERINE/THREONINE PROTEIN KINASE, TRANSFERASE
1478	1dm		2	73	2.1e-12	0.59	0.88		DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE; CHAIN: NULL;	TRANSFERASE ALPHA BETA FOLD
1478	1euw	A	2	73	2.2e-10	0.36	0.11		DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE; CHAIN: A;	HYDROLASE DUTPASE, DUTP PYROPHOSPHATASE; HYDROLASE, DUTPASE, ELAV, TRIMERIC ENZYME, ASPARTYL PROTEASE
1478	1f7d	A	2	73	1.7e-12	0.07	0.95		POL POLYPYRROLINE 5'-NUCLEOTIDOHYDROLASE; CHAIN: A;	HYDROLASE DUTPASE, JELLY ROLL, MERCURY DERIVATIVE
1478	1f7r	A	2	73	1.3e-12	0.36	0.98		POL POLYPYRROLINE; CHAIN: A, B;	VIRUS/VIRAL PROTEIN EIGHT STRANDED BETA-BARREL
1479	1bj8		10	113	1.6e-19	0.72	0.77		GP130; CHAIN: NULL;	VIRUS/VIRAL PROTEIN EIGHT STRANDED BETA BARREL PROTEIN
										RECEPTOR RECEPTOR, SIGNAL TRANSDUCER OF IL-6 TYPE CYTOKINES, THIRD 2 N-TERMINAL DOMAIN, TRANSMEMBRANE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1479	1bpv		9	113	3.2e-18			54.68	TITIN; CHAIN: NULL;	GLYCOPROTEIN CONNECTIN A71, CONNECTIN; TITIN, CONNECTIN; FIBRONECTIN TYPE III
1479	1bqu	A	6	115	1.1e-20	0.45	-0.06		GP130; CHAIN: A, B;	SIGNALING PROTEIN CYTOKINE RECEPTOR, GLYCOPROTEIN 130, GP130, INTERLEUKINE 6 2 RECEPTOR BETA SUBUNIT, SIGNALING PROTEIN
1479	1cb		10	188	4.8e-26	0.01	-0.13		NEURAL ADHESION MOLECULE DROSOPHILA NEUROGLIAN (CHYMOTRYPTIC FRAGMENT CONTAINING THE ICB 3 TWO AMINO PROXIMAL FIBRONECTIN TYPE III REPEATS ICB 4 (RESIDUES 610 - 814)) ICFB 5	
1479	1fhh	A	1	185	3.2e-25	0.01	-0.11		FIBRONECTIN; CHAIN: A;	HEPARIN AND INTEGRIN BINDING HEPARIN AND INTEGRIN BINDING
1479	1qg3	A	15	187	9.6e-23	0.18	-0.09		INTEGRIN BETA 4 SUBUNIT; CHAIN: A, B;	STRUCTURAL PROTEIN INTEGRIN, HEMIDESMOSOME, FIBRONECTIN, CARCINOMA, STRUCTURAL 2 PROTEIN
1480	1abt	A	14	51	0.00013	-0.67	0.13		TOXIN ALPHA-	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
									BUNGAROTOXIN COMPLEXED WITH THE 185 - 196 FRAGMENT OF 1ABT 3 THE ALPHA-SUBUNIT OF THE TORPEDO NICOTINIC ACETYLCHOLINE 1ABT 4 RECEPTOR (NMR, 4 STRUCTURES) 1ABT 5 POSTSYNAPTIC NEUROTOXIN ALPHA-*BUNGAROTOXIN 2ABX 4 CARDIOTOXIN CARDIOTOXIN CTX 1 (NMR, 11 STRUCTURES) 2CDX 3	
1480	2abx	A	14	51	4.3e-05	-0.67	0.18			
1480	2cdx		15	51	0.0086	-0.70	0.04			
1482	2lbp		55	188	8e-19	-0.04	0.01		PERIPLASMIC BINDING PROTEIN (LBPS) 2LBP 4	
1482	2liv		54	189	4.8e-18	0.01	0.09		PERIPLASMIC BINDING PROTEIN LEUCINE-BINDING LEUCINE(SLASH)*ISOLEUCIN E(SLASH)*VALINE-BINDING PROTEIN 2LIV 4 (LIVBPS) 2LIV 5	
1485	1cot	A	87	128	8.6e-05	-0.69	0.16		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1485	1E21	A	51	143	1.6e-09	-0.29	0.03		TRANSCRIPTASE (B-CHAIN); CHAIN: B;	DRUG DESIGN
1485	1Hrh	A	52	157	0.00013	-0.25	0.52		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H DOMAIN OF HIV-1 $\beta$ REVERSE TRANSCRIPTASE 1HRH 3	HYDROLASE RNASE H, NUCLEASE, RNASE H*, RIBNUCLEASE H, METAL-BINDING 2 PROTEIN, PROTEIN FOLDING
1485	1Rtl		48	157	4.8e-09	0.02	0.05		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H (E.C.3.1.26.4) 1RTL 3	
1485	1Rtl		87	163	8.6e-06	0.05	0.15		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H (E.C.3.1.26.4) 1RTL 3	
1485	1Rth	A	22	157	9.6e-05	-0.16	0.36		HIV-1 REVERSE TRANSCRIPTASE, 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT, 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1485	1Vrt	A	86	157	0.00032	-0.31	0.00		HIV-1 REVERSE TRANSCRIPTASE, 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT, 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
1506	1bal		1	315	3.2e-99			139.00	HEAT-SHOCK COGNATE 70KD PROTEIN; CHAIN: NULL;	HYDROLASE HYDROLASE, ACTING ON ACID

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1506	1ba1		12	314	3.2e-99	0.21	1.00		HEAT-SHOCK COGNATE 70KD PROTEIN; CHAIN: NULL;	ANHYDRIDES, ATP-BINDING, 2 HEAT SHOCK
1506	1bpr		233	405	1.7e-51			86.63	DNAA; CHAIN: NULL;	MOLECULAR CHAPERONE MOLECULAR CHAPERONE, HSP70, PEPTIDE BINDING, PROTEIN 2 FOLDING
1506	1bpr		271	400	1.7e-51	-0.42	0.96		DNAA; CHAIN: NULL;	MOLECULAR CHAPERONE MOLECULAR CHAPERONE, HSP70, PEPTIDE BINDING, PROTEIN 2 FOLDING
1506	1ckr	A	237	391	8.6e-45			137.48	HEAT SHOCK SUBSTRATE BINDING DOMAIN OF HSC-70; CHAIN: A;	CHAPERONE MOLECULAR CHAPERONE, HSP70, PEPTIDE BINDING, PROTEIN FOLDING
1506	1ckr	A	271	391	4.8e-28	0.04	1.00		HEAT SHOCK SUBSTRATE BINDING DOMAIN OF HSC-70; CHAIN: A;	CHAPERONE MOLECULAR CHAPERONE, HSP70, PEPTIDE BINDING, PROTEIN FOLDING
1506	1ckr	A	277	391	8.6e-45	0.03	1.00		HEAT SHOCK SUBSTRATE BINDING DOMAIN OF HSC-70; CHAIN: A;	CHAPERONE MOLECULAR CHAPERONE, HSP70, PEPTIDE BINDING, PROTEIN FOLDING
1506	1dg4	A	271	357	1.7e-35	0.09	0.71		DNAA; CHAIN: A;	CHAPERONE DNAA, CHAPERONE, SUBSTRATE BINDING DOMAIN
1506	1dkg	D	1	347	3.2e-81			103.04	NUCLEOTIDE EXCHANGE FACTOR GRPE; CHAIN: A, B;	COMPLEX (HSP24/HSP70) HSP70, GRPE, MOLECULAR

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1506	1dkg	D	82	314	3.2e-81	-0.07	0.87		MOLECULAR CHAPERONE D <sub>NAK</sub> ; CHAIN: D;	CHAPERONE, NUCLEOTIDE EXCHANGE 2 FACTOR, COILED-COIL, COMPLEX (HSP24/HSP70)
1506	1dkx	A	241	424	1.7e-52			70.76	NUCLEOTIDE EXCHANGE FACTOR GRPE; CHAIN: A, B; MOLECULAR CHAPERONE D <sub>NAK</sub> ; CHAIN: D;	COMPLEX (HSP24/HSP70) HSP70, GRPE, MOLECULAR CHAPERONE, NUCLEOTIDE EXCHANGE 2 FACTOR, COILED-COIL, COMPLEX (HSP24/HSP70)
1506	1dkx	A	271	424	1.7e-52	0.15	1.00		SUBSTRATE BINDING DOMAIN OF D <sub>NAK</sub> ; CHAIN: A; SUBSTRATE PEPTIDE (7 RESIDUES); CHAIN: B;	COMPLEX (MOLECULAR CHAPERONE/PEPTIDE) D <sub>NAK</sub> , HEAT SHOCK PROTEIN 70 KDA (HSP70), COMPLEX 2 (MOLECULAR CHAPERONE/PEPTIDE)
1506	1dkx	A	272	422	6.4e-24	-0.20	1.00		SUBSTRATE BINDING DOMAIN OF D <sub>NAK</sub> ; CHAIN: A; SUBSTRATE PEPTIDE (7 RESIDUES); CHAIN: B;	COMPLEX (MOLECULAR CHAPERONE/PEPTIDE) D <sub>NAK</sub> , HEAT SHOCK PROTEIN 70 KDA (HSP70), COMPLEX 2 (MOLECULAR CHAPERONE/PEPTIDE)
1506	1dky	B	241	424	2.6e-53			80.32	D <sub>NAK</sub> ; CHAIN: A, B; PEPTIDE SUBSTRATE; CHAIN: C, D;	COMPLEX (MOLECULAR CHAPERONE/PEPTIDE) D <sub>NAK</sub> ,



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1506	1dky	B	271	424	2.6e-53	0.16	1.00		DNAK; CHAIN: A, B; PEPTIDE SUBSTRATE; CHAIN: C, D;	HEAT SHOCK PROTEIN 70 KDA (HSP70), COMPLEX 2 (MOLECULAR CHAPERONE/PEPTIDE)
1506	1dky	B	272	422	6.4e-24	0.10	1.00		DNAK; CHAIN: A, B; PEPTIDE SUBSTRATE; CHAIN: C, D;	COMPLEX (MOLECULAR CHAPERONE/PEPTIDE) DNAK, CHAPERONE/PEPTIDE) DNAK, HEAT SHOCK PROTEIN 70 KDA (HSP70), COMPLEX 2 (MOLECULAR CHAPERONE/PEPTIDE)
1506	1hjo	A	1	316	4.8e-98			120.78	HEAT-SHOCK 70KD PROTEIN; CHAIN: A;	HYDROLASE ATP-BINDING, CHAPERONE, HEAT SHOCK, HYDROLASE
1506	1hjo	A	82	316	4.8e-98	0.03	1.00		HEAT-SHOCK 70KD PROTEIN; CHAIN: A;	HYDROLASE ATP-BINDING, CHAPERONE, HEAT SHOCK, HYDROLASE
1515	1qf6	A	16	241	1.6e-68	0.42	0.99		THREONYL-TRNA SYNTHETASE; CHAIN: A; THREONINE TRNA; CHAIN: B;	LIGASE/RNA THRS; TRNA (THR); THREONYL-TRNA SYNTHETASE, TRNA(THR), AMP, ZINC, MRNA, 2 AMINOACYLATION,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Pst Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1530	1zt1	A	65	155	3.2e-18	-0.22	0.10		RIBONUCLEASE H; CHAIN: A;	TRANSLATIONAL REGULATION, PROTEIN/RNA
1530	1hrh	A	58	149	9.6e-16	-0.34	0.04		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H DOMAIN OF /HIV-1\$ REVERSE TRANSCRIPTASE 1HRH 3	HYDROLASE RNASE H, NUCLEASE, RNASE H*, RIBNUCLEASE H, METAL-BINDING 2 PROTEIN, PROTEIN FOLDING
1530	1rll		65	155	9.6e-15	-0.10	0.31		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H (E.C.3.1.26.4) 1RLL 3	
1537	1dt6	A	31	153	1.8e-21	-0.22	1.00		CYTOKROME P450 2C5; CHAIN: A;	OXIDOREDUCTASE PROGESTERONE 21-HYDROXYLASE, CYP11C5 P450 1, MEMBRANE PROTEIN, PROGESTERONE 21-HYDROXYLASE, BENZO(A) 2 PYRENE HYDROXYLASE, ESTRADIOL 2-HYDROXYLASE, P450, CYP2C5
1564	1a0p		11	206	1.4e-37	0.08	-0.02		SITE-SPECIFIC RECOMBINASE XERD; CHAIN: NULL;	DNA RECOMBINATION XERD, RECOMBINASE, DNA BINDING,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1564	1ae9	A	43	210	3.2e-16	0.07	-0.07		LAMBDA INTEGRASE; CHAIN: A, B;	DNA RECOMBINATION DNA RECOMBINATION DNA RECOMBINATION, INTEGRASE, SITE-SPECIFIC RECOMBINATION
1564	1ae9	B	43	193	8e-17	0.09	0.12		LAMBDA INTEGRASE; CHAIN: A, B;	DNA RECOMBINATION DNA RECOMBINATION, INTEGRASE, SITE-SPECIFIC RECOMBINATION
1564	1aih	A	36	212	6.4e-26	0.23	-0.02		HPI INTEGRASE; CHAIN: A, B, C, D;	DNA INTEGRATION DNA INTEGRATION, RECOMBINATION
1566	1bqg		54	339	8e-31	0.22	0.58		D-GLUCARATE DEHYDRATASE; CHAIN: NULL;	GLUCARATE GLUCARATE, TIM BARREL, ENOLASE SUPERFAMILY
1566	1chr	A	21	352	3.2e-51			61.26	ISOMERASE CHLOROMUCONATE CYCLOISOMERASE (E.C.5.1.7) ICHR 3	
1566	1chr	A	59	350	3.2e-51	0.44	0.68		ISOMERASE CHLOROMUCONATE CYCLOISOMERASE (E.C.5.1.7) ICHR 3	
1566	1ec7	A	62	339	1.3e-27	0.07	0.54		GLUCARATE DEHYDRATASE; CHAIN: A, B, C, D;	LYASE GLUCARATE DEHYDRATASE ENOLASE ENZYME SUPERFAMILY TIM BARREL 2 (BETA/ALPHA)/BETA

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Pst Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1566	1flu	A	165	337	9.6e-19	0.20	-0.05		O-SUCCINYLBENZOATE SYNTHASE; CHAIN: A;	BARREL OXIDOREDUCTASE ENOLASE SUPERFAMILY
1566	1thv	A	138	337	3.2e-21	0.23	0.31		O-SUCCINYLBENZOATE SYNTHASE; CHAIN: A;	OXIDOREDUCTASE ENOLASE SUPERFAMILY
1566	1mdl		159	351	9e-34	0.40	1.00		MANDELATE RACEMASE; CHAIN: NULL;	ISOMERASE ISOMERASE, MANDELATE PATHWAY, MAGNESIUM
1566	1mdl		5	353	3.2e-51			66.01	MANDELATE RACEMASE; CHAIN: NULL;	ISOMERASE ISOMERASE, MANDELATE PATHWAY, MAGNESIUM
1566	1mdl		60	337	3.2e-51	0.35	0.86		MANDELATE RACEMASE; CHAIN: NULL;	ISOMERASE ISOMERASE, MANDELATE PATHWAY, MAGNESIUM
1566	1muc	A	60	351	9.6e-51	0.30	0.25		MUCONATE LACTONIZING ENZYME; CHAIN: A, B;	ISOMERASE CIC, CIS MUCONATE CYCLOISOMERASE; MUCONATE LACTONIZING ENZYME
1566	1one	A	40	323	4.8e-85	-0.24	0.06		ENOLASE; CHAIN: A, B;	LYASE 2-PHOSPHO-D-GLYCERATE HYDROLASE; LYASE, GLYCOLYSIS
1570	1aft		68	138	8e-23	0.44	0.22		MERP; CHAIN: NULL;	MERCURY DETOXIFICATION MERCURIC TRANSPORT PROTEIN; MERCURY DETOXIFICATION, PERIPLASMIC, HEAVY METAL

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1570	1aw0		68	138	3.2e-19	0.41	0.64			TRANSPORT, 2 ALPHA-BETA SANDWICH
										HYDROLASE COPPER-TRANSPORTING ATPASE, COPPER-BINDING DOMAIN, HYDROLASE
1570	1cc8	A	66	135	6.4e-10	0.13	0.22		METALLOCHAPERONE ATX1; CHAIN: A;	METAL TRANSPORT COPPER TRANSPORT, MERCURY COORDINATION, METAL TRANSPORT
1570	1cpz	A	72	136	1.3e-18	0.67	0.42		COPZ; CHAIN: A;	GENE REGULATION COPPER CHAPERONE, METAL TRANSPORT, GENE REGULATION
1592	1aab		126	192	0.0016	-0.17	0.10		HIGH MOBILITY GROUP PROTEIN; 1AAB 5 CHAIN: NULL; 1AAB 6	DNA-BINDING HMGA DNA-BINDING HMG-BOX DOMAIN A OF RAT HMGI; 1AAB 8 HMG-BOX 1AAB 20
1592	1aab		160	187	0.00045	-0.62	0.34		HIGH MOBILITY GROUP PROTEIN; 1AAB 5 CHAIN: NULL; 1AAB 6	DNA-BINDING HMGA DNA-BINDING HMG-BOX DOMAIN A OF RAT HMGI; 1AAB 8 HMG-BOX 1AAB 20
1592	1cg7	A	126	188	9.6e-08	-0.34	0.00		NON HISTONE PROTEIN 6 A; CHAIN: A;	DNA BINDING PROTEIN HMGBOX, DNA BENDING, DNA RECOGNITION, CHROMATIN, NMR, DNA 2 BINDING PROTEIN
1592	1ckt	A	133	192	0.0062	-0.07	0.25		HIGH MOBILITY GROUP 1	GENE REGULATION/DNA HMGB-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
									PROTEIN; CHAIN: A; DNA (5'-D(*CP*CP*(DDO) CHAIN: B; DNA (5'-CHAIN: C;	1, AMPHOTERIN, HEPARIN-BINDING PROTEIN P30, HIGH-MOBILITY GROUP DOMAIN, BENT DNA, PROTEIN-DRUG-DNA 2 COMPLEX, GENE REGULATION/DNA
1592	1c1t	A	160	187	0.0018	-0.73	0.74		HIGH MOBILITY GROUP 1 PROTEIN; CHAIN: A; DNA (5'-D(*CP*CP*(DDO) CHAIN: B; DNA (5'-CHAIN: C;	GENE REGULATION/DNA HMG-1, AMPHOTERIN, HEPARIN-BINDING PROTEIN P30, HIGH-MOBILITY GROUP DOMAIN, BENT DNA, PROTEIN-DRUG-DNA 2 COMPLEX, GENE REGULATION/DNA
1592	1hme		133	187	3.2e-06	-0.42	0.17		DNA-BINDING HIGH MOBILITY GROUP PROTEIN FRAGMENT-B (HMG1) (DNA-BINDING IHME 3 HMG-BOX DOMAIN B OF RAT HMG1) (NMR, 1 STRUCTURE) IHME 4	
1592	1hsm		133	187	3.2e-06	0.00	0.31		DNA-BINDING HIGH MOBILITY GROUP PROTEIN 1 (HMG1) BOX 2, COMPLEXED WITH IHSM 3 MERCAPTOETHANOL (NMR, MINIMIZED AVERAGE STRUCTURE) IHSM 4	
1592	2lef	A	131	207	3.2e-19	-0.13	0.04		LYMPHOID ENHANCER-BINDING FACTOR; CHAIN: A;	GENE REGULATION/DNA LEF-1 HMG; LEFT1, HMG, TCR-A;

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									DNA (5'-CHAIN: B; DNA (5'-CHAIN: C;	TRANSCRIPTION FACTOR, DNA BINDING, DNA 2 BENDING, COMPLEX (HMG DOMAIN/DNA), GENE REGULATION/DNA
1601	1czz	A	7	238	8e-05	0.04	0.00		TOLB PROTEIN; CHAIN: A;	TOXIN BINDING PROTEIN TWO DOMAINS: BETA PROPELLER AND ALPHA/BETA FOLD
1601	1etj	A	8	258	1.6e-66	0.37	1.00		TRANSCRIPTIONAL REPRESSOR TUP1; CHAIN: A, B, C;	TRANSCRIPTION INHIBITOR BETA-PROPELLER
1601	1got	B	5	256	4.8e-71	0.28	0.88		GT-ALPHA/GT-ALPHA CHIMERA; CHAIN: A; GT-BETA; CHAIN: B; GT-GAMMA; CHAIN: G;	COMPLEX (GTP-BINDING/TRANSDUCER) BETA1, TRANSDUCIN BETA SUBUNIT, GAMMA1, TRANSDUCIN GAMMA SUBUNIT; COMPLEX (GTP-BINDING/TRANSDUCER), G PROTEIN, HETEROTRIMER 2 SIGNAL TRANSDUCTION
1610	1ako		232	334	4.5e-14	0.07	0.21		EXONUCLEASE III; CHAIN: NULL;	NUCLEASE NUCLEASE, EXONUCLEASE, AP- ENDONUCLEASE, DNA REPAIR
1610	1bix		209	333	3.2e-05	0.46	0.57		AP ENDONUCLEASE I; CHAIN: NULL;	DNA REPAIR DNA REPAIR, ENDONUCLEASE, HAP1, REF-1, ABASIC SITE 2 RECOGNITION
1612	1ait	A	158	223	3.2e-14	0.39	0.24		NUCLEOCAPSID PROTEIN;	COMPLEX (NUCLEOCAPSID

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN: A; SL3 STEM-LOOP RNA; CHAIN: B;	PROTEIN/RNA) NUCLEOCAPSID PROTEIN, COMPLEX (NUCLEOCAPSID PROTEIN/RNA), 2 STEM-LOOP RNA
1612	1aaf		158	223	3.2e-14	0.17	0.12		NUCLEOCAPSID PROTEIN HIV-1 NUCLEOCAPSID PROTEIN (MN ISOLATE) (NMR, 20 STRUCTURES) 1AAF	
1612	1bj6	A	166	221	3.2e-12	0.37	0.15		DNA (ACGCC); CHAIN: D; NUCLEOCAPSID PROTEIN 7; CHAIN: A;	COMPLEX (NUCLEOCAPSID PROTEIN/DNA) (12-53)NCP7; COMPLEX (NUCLEOCAPSID PROTEIN/DNA), NUCLEIC ACID, 2 RETROVIRUS, VIRUS MORPHOGENESIS, ZINC FINGER
1612	1bj6	A	203	239	1.4e-08	0.16	0.06		DNA (ACGCC); CHAIN: D; NUCLEOCAPSID PROTEIN 7; CHAIN: A;	COMPLEX (NUCLEOCAPSID PROTEIN/DNA) (12-53)NCP7; COMPLEX (NUCLEOCAPSID PROTEIN/DNA), NUCLEIC ACID, 2 RETROVIRUS, VIRUS MORPHOGENESIS, ZINC FINGER
1612	1cl4	A	201	228	2.7e-09	0.64	0.88		GAG POLYPEPTIDE; CHAIN: A;	VIRAL PROTEIN NUCLEOCAPSID PROTEIN, RNA BINDING PROTEIN, RETROVIRUS, 2 VIRAL PROTEIN
1612	1dsv	A	200	229	4.8e-05	0.39	0.99		NUCLEIC ACID BINDING PROTEIN P14; CHAIN: A;	VIRUS/VIRAL PROTEIN CCHC TYPE ZINC FINGER,



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1612	1dsv	A	201	228	4.5e-10	0.25	1.00		NUCLEIC ACID BINDING PROTEIN P14; CHAIN: A;	VIRUS/VIRAL PROTEIN VIRUS/VIRAL PROTEIN CCHC TYPE ZINC FINGER, VIRUS/VIRAL PROTEIN
1616	1iam		19	128	8e-36	-0.09	0.74		INTERCELLULAR ADHESION MOLECULE-1; CHAIN: NULL;	RHINOVIRUS RECEPTOR ICAM-1, CD54; RHINOVIRUS RECEPTOR, CELL ADHESION, INTEGRIN LIGAND, 2 GLYCOPROTEIN, LFA-1 LIGAND, IMMUNOGLOBULIN FOLD, 3 TRANSMEMBRANE
1616	1zsq		32	135	6.4e-33	0.19	1.00		INTERCELLULAR ADHESION MOLECULE-2; CHAIN: NULL;	CELL ADHESION ICAM-2; IMMUNOGLOBULIN FOLD, CELL ADHESION, GLYCOPROTEIN, 2 TRANSMEMBRANE, REPEAT, SIGNAL
1618	1aub		139	183	8e-12	-0.28	0.31		HIV-2 INTEGRASE; CHAIN: NULL;	INTEGRASE INTEGRASE, AIDS, POLYPROTEIN
1618	1z21	A	25	153	8e-21	-0.09	0.17		RIBONUCLEASE HI; CHAIN: A;	HYDROLASE RNASE H, NUCLEASE, RNASE H*, RIBONUCLEASE H, METAL-BINDING 2 PROTEIN, PROTEIN FOLDING
1618	1lth	A	20	120	1.3e-24	-0.19	0.03		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H	

Table 5

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Pst Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1641	1a7a	A	42	86	0.00048	0.55	1.00		S-ADENOSYLHOMOCYSTEINE HYDROLASE; CHAIN: A, B;	HYDROLASE HYDROLASE, NAD BINDING PROTEIN
1641	1ae1	A	39	297	9.6e-64			79.01	TROPINONE REDUCTASE-I; CHAIN: A, B;	OXIDOREDUCTASE OXIDOREDUCTASE, TROPANE ALKALOID BIOSYNTHESIS, REDUCTION OF 2 TROPINONE TO TROPINE, SHORT-CHAIN DEHYDROGENASE
1641	1ae1	A	39	313	9.6e-64	0.27	1.00		TROPINONE REDUCTASE-I; CHAIN: A, B;	OXIDOREDUCTASE OXIDOREDUCTASE, TROPANE ALKALOID BIOSYNTHESIS, REDUCTION OF 2 TROPINONE TO TROPINE, SHORT-CHAIN DEHYDROGENASE
1641	1ae1	B	39	312	3.2e-66			74.73	TROPINONE REDUCTASE-I; CHAIN: A, B;	OXIDOREDUCTASE OXIDOREDUCTASE, TROPANE ALKALOID BIOSYNTHESIS, REDUCTION OF 2 TROPINONE TO TROPINE, SHORT-CHAIN DEHYDROGENASE
1641	1ae1	B	39	313	3.2e-66	0.37	1.00		TROPINONE REDUCTASE-I; CHAIN: A, B;	OXIDOREDUCTASE OXIDOREDUCTASE, TROPANE ALKALOID BIOSYNTHESIS, REDUCTION OF 2 TROPINONE TO TROPINE, SHORT-CHAIN DEHYDROGENASE
1641	1b16	A	40	255	3.2e-21	0.40	1.00		ALCOHOL DEHYDROGENASE;	OXIDOREDUCTASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
									CHAIN: A, B;	OXIDOREDUCTASE, DETOXIFICATION, METABOLISM, ALCOHOL 2 DEHYDROGENASE, DROSOPHILA LEBANONENSIS, SHORT-CHAIN 3 DEHYDROGENASES/REDUCTASES, TERNARY COMPLEX, NAD-3-PENTANONE 4 ADDUCT
1641	1bdb		40	324	1.6e-45			69.17	CIS-BIPHENYL-2,3-DIHYDRODIOL-2,3-DEHYDROGENASE; CHAIN: NULL;	OXIDOREDUCTASE NAD-DEPENDENT OXIDOREDUCTASE, SHORT-CHAIN ALCOHOL 2 DEHYDROGENASE, PCB DEGRADATION
1641	1bdb		41	314	1.6e-45	0.24	1.00		CIS-BIPHENYL-2,3-DIHYDRODIOL-2,3-DEHYDROGENASE; CHAIN: NULL;	OXIDOREDUCTASE NAD-DEPENDENT OXIDOREDUCTASE, SHORT-CHAIN ALCOHOL 2 DEHYDROGENASE, PCB DEGRADATION
1641	1cyd	A	40	311	8e-53			79.57	CARBONYL REDUCTASE; CHAIN: A, B, C, D;	OXIDOREDUCTASE SHORT-CHAIN DEHYDROGENASE, OXIDOREDUCTASE
1641	1cyd	A	41	312	8e-53	0.47	1.00		CARBONYL REDUCTASE; CHAIN: A, B, C, D;	OXIDOREDUCTASE SHORT-CHAIN DEHYDROGENASE, OXIDOREDUCTASE
1641	1db3	A	44	258	4e-09	0.05	0.76		GDP-MANNOSE 4,6-	LYASE DEHYDRATASE, NADP,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1641	1dhr		43	256	9e-24	0.10	0.36		DEHYDRATASE; CHAIN: A; OXIDOREDUCTASE(ACTING ON NADH OR NADPH) DIHYDROPTERIDINE REDUCTASE (DHPR) (E.C.1.6.99.10) COMPLEX 1DHR 3 WITH NADH 1DHR 4	GDP-MANNNOSE, GDP-FUCOSE
1641	1ek6	A	44	242	3.6e-10	0.48	1.00		UDP-GALACTOSE 4- EPIMERASE; CHAIN: A, B;	ISOMERASE EPIMERASE, SHORT-CHAIN DEHYDROGENASE, GALACTOSEMIA
1641	1eny		38	312	1.3e-29			53.65	ENOYL-ACYL CARRIER PROTEIN (ACP) REDUCTASE; 1ENY 4 CHAIN: NULL; 1ENY 5	OXIDOREDUCTASE INHA; 1ENY 6
1641	1eny		39	258	1.3e-29	0.29	0.94		ENOYL-ACYL CARRIER PROTEIN (ACP) REDUCTASE; 1ENY 4 CHAIN: NULL; 1ENY 5	OXIDOREDUCTASE INHA; 1ENY 6
1641	1eq2	A	46	263	9e-09	-0.19	0.49		ADP-L-GLYCERO-D- MANNNOHEPTOSE 6- EPIMERASE; CHAIN: A, B, C, D, E, F, G, H, I, J;	ISOMERASE N-TERMINAL DOMAIN ROSSMANN FOLD, C- TERMINAL MIXED 2 ALPHA/BETA DOMAIN; SHORT- CHAIN DEHYDROGENASE/REDUCTASE FOLD
1641	1fbs		42	335	2.7e-31			67.54	17-BETA-HYDROXYSTEROID- DEHYDROGENASE; CHAIN: NULL;	DEHYDROGENASE DEHYDROGENASE, 17-BETA- HYDROXYSTEROID
1641	1fbs		43	256	2.7e-31	0.29	1.00		17-BETA-HYDROXYSTEROID- NULL;	DEHYDROGENASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
									DEHYDROGENASE; CHAIN: NULL;	DEHYDROGENASE, 17-BETA-HYDROXYSTEROID
1641	1fda		45	299	1.4e-26	0.36	0.89		17-BETA-HYDROXYSTEROID-DEHYDROGENASE; CHAIN: NULL;	DEHYDROGENASE DEHYDROGENASE, 17-BETA-HYDROXYSTEROID
1641	1fmc	A	35	310	3.2e-67			84.94	7 ALPHA-HYDROXYSTEROID DEHYDROGENASE; CHAIN: A, B;	OXIDOREDUCTASE SHORT-CHAIN DEHYDROGENASE/REDUCTASE , BILE ACID CATABOLISM
1641	1fmc	A	40	308	3.2e-67	0.40	1.00		7 ALPHA-HYDROXYSTEROID DEHYDROGENASE; CHAIN: A, B;	OXIDOREDUCTASE SHORT-CHAIN DEHYDROGENASE/REDUCTASE , BILE ACID CATABOLISM
1641	1hdc	A	39	313	6.4e-63	0.50	1.00		OXIDOREDUCTASE 3-ALPHA, 20-BETA-HYDROXYSTEROID DEHYDROGENASE (E.C.1.1.1.53) 1HDC 3 COMPLEXED WITH CARBENOXOLONE 1HDC 4	
1641	1hdc	A	39	318	6.4e-63			69.01	OXIDOREDUCTASE 3-ALPHA, 20-BETA-HYDROXYSTEROID DEHYDROGENASE (E.C.1.1.1.53) 1HDC 3 COMPLEXED WITH CARBENOXOLONE 1HDC 4	
1641	1leh	A	40	116	0.0022	0.42	0.24		LEUCINE DEHYDROGENASE; CHAIN: A, B;	OXIDOREDUCTASE OXIDOREDUCTASE
1641	1oaa		38	304	2.7e-30			52.72	SEPIAPTERIN REDUCTASE;	OXIDOREDUCTASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
									CHAIN: NULL;	SEPIAPTERIN REDUCTASE, TETRAHYDROBIOPTERIN, OXIDOREDUCTASE
1641	1oa2		43	291	2.7e-30	0.30	0.95		SEPIAPTERIN REDUCTASE; CHAIN: NULL;	OXIDOREDUCTASE SEPIAPTERIN REDUCTASE, TETRAHYDROBIOPTERIN, OXIDOREDUCTASE
1641	1oa2		45	267	3.2e-17	0.21	1.00		SEPIAPTERIN REDUCTASE; CHAIN: NULL;	OXIDOREDUCTASE SEPIAPTERIN REDUCTASE, TETRAHYDROBIOPTERIN, OXIDOREDUCTASE
1641	1qtr	A	44	248	1.4e-08	0.27	0.87		SULFOLIPID BIOSYNTHESIS (SQD1) PROTEIN; CHAIN: A;	ISOMERASE ROSSMANN FOLD, SHORT HYDROGEN BONDS, SDR HOMOLOG, ISOMERASE
1641	1udb		9	336	4.5e-08			56.43	UDP-GALACTOSE-4-EPIMERASE; CHAIN: NULL;	ISOMERASE EPIMERASE; UDP-GALACTOSE, EPIMERASE, ISOMERASE
1641	1ybv	A	22	307	8e-61			84.54	TRIHYDROXYNAPHTHALENE REDUCTASE; CHAIN: A, B;	OXIDOREDUCTASE NAPHTHOL REDUCTASE;
1641	1ybv	A	35	311	8e-61	0.56	1.00		TRIHYDROXYNAPHTHALENE REDUCTASE; CHAIN: A, B;	OXIDOREDUCTASE NAPHTHOL REDUCTASE;
1641	2ac2	A	36	307	1.6e-63			71.59	TROPINONE REDUCTASE-II; CHAIN: A, B;	OXIDOREDUCTASE OXIDOREDUCTASE, TROPANE ALKALOID BIOSYNTHESIS, REDUCTION OF 2 TROPINONE TO PSEUDOTROPINE, SHORT-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1641	2ac2	A	39	312	1.6e-63	0.10	1.00		TROPINONE REDUCTASE-II; CHAIN: A, B;	CHAIN DEHYDROGENASE OXIDOREDUCTASE OXIDOREDUCTASE, TROPANE ALKALOID BIOSYNTHESIS, REDUCTION OF 2 TROPINONE TO PSEUDOTROPINE, SHORT-CHAIN DEHYDROGENASE
1641	2dlid	A	41	96	0.00018	0.47	0.31		D-LACTATE DEHYDROGENASE; 2DLD 5 CHAIN: A, B; 2DLD 6	OXIDOREDUCTASE (CHOH(D)-NAD(+)) R-LACTATE DEHYDROGENASE; 2DLD 7
1641	3hda	C	48	107	0.00096	0.26	0.03		L-3-HYDROXYACYL COA DEHYDROGENASE; CHAIN: A, B, C;	OXIDOREDUCTASE SCHAD; OXIDOREDUCTASE, BETA OXIDATION, SCHAD, CATALYTIC ACTIVITY: 2 L-3-HYDROXYACYL-COA + NAD(+) = 3-OXOACYL-COA + NADH
1667	1a3r	L	46	220	3.2e-59	0.23	-0.13		IGG2A; CHAIN: L, H; HUMAN RHINOVIRUS CAPSID PROTEIN VP2; CHAIN: P;	COMPLEX (IMMUNOGLOBULIN/VRAL PEPTIDE) ANTIBODY 8F5; IMMUNOGLOBULIN, ANTIBODY, RHINOVIRUS, NEUTRALIZATION, 2 CONTINUOUS EPITOPE, COMPLEX (IMMUNOGLOBULIN/VRAL PEPTIDE)
1667	1a4f	L	46	270	1.3e-50			50.13	IMMUNOGLOBULIN, DIEIS	IMMUNOGLOBULIN



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1667	1axt	L	46	283	1.4e-51			50.38	ALDER CATALYTIC ANTIBODY; CHAIN: L, H, A, B;	IMMUNOGLOBULIN, ANTIBODY, CATALYTIC ANTIBODY, DIELS ALDER, 2 GERMLINE
1667	1b2w	L	46	205	1.6e-56	0.06	-0.15		ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H;	IMMUNE SYSTEM IMMUNOGLOBULIN; IMMUNOGLOBULIN ANTIBODY ENGINEERING, HUMANIZED AND CHIMERIC ANTIBODY, FAB, 2 X-RAY STRUCTURE, THREE-DIMENSIONAL STRUCTURE, GAMMA-3 INTERFERON, IMMUNE SYSTEM
1667	1bj1	L	46	205	1.1e-57	0.12	-0.11		FAB FRAGMENT; CHAIN: L, H, J, K; VASCULAR ENDOTHELIAL GROWTH FACTOR; CHAIN: V, W;	COMPLEX (ANTIBODY/ANTIGEN) FAB-12; VEGF; COMPLEX (ANTIBODY/ANTIGEN), ANGIOGENIC FACTOR
1667	1c5c	H	164	359	3.2e-45	0.08	-0.20		CHIMERIC DECARBOXYLASE ANTIBODY 21D8; CHAIN: L; CHIMERIC DECARBOXYLASE ANTIBODY 21D8; CHAIN: H;	IMMUNE SYSTEM IMMUNOGLOBULIN, CATALYTIC ANTIBODY, CHIMERIC FAB, 2 DECARBOXYLASE, HAPTEN COMPLEX

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1667	1clz	L	46	286	3.2e-52			50.34	IGG FAB (GGG3, KAPPA); CHAIN: L, H;	IMMUNOGLOBULIN MER96 FAB (IMMUNOGLOBULIN); IMMUNOGLOBULIN C REGION, GLYCOPROTEIN, TRANSMEMBRANE
1667	1dbb	L	46	283	1.6e-52			54.63	IMMUNOGLOBULIN FAB <sup>1</sup> FRAGMENT OF THE DB3 ANTI-STERIOD MONOCLONAL ANTIBODY 1DBB 3 (GGG1, SUBGROUP 2A, KAPPA 1) COMPLEX WITH PROGESTERONE 1DBB 4	
1667	1dee	A	46	205	1.6e-58	0.09	-0.09		IGM RF 2A2; CHAIN: A, C, E; IGM RF 2A2; CHAIN: B, D, F; IMMUNOGLOBULIN G BINDING PROTEIN A; CHAIN: G, H;	IMMUNE SYSTEM FAB-IBP COMPLEX CRYSTAL STRUCTURE 2.7A RESOLUTION BINDING 2 OUTSIDE THE ANTIGEN COMBINING SITE SUPERANTIGEN FAB VH3 3 SPECIFICITY
1667	1dzb	A	46	286	1.1e-85	0.46	-0.15		SCFV FRAGMENT 1F9; CHAIN: A, B; TURKEY EGG-WHITE LYSOZYME C; CHAIN: X, Y;	COMPLEX (ANTIBODY ANTIGEN) 1.4BETA-N-ACETYLMURAMIDASE C <sub>1</sub> SINGLE-DOMAIN ANTIBODY, TURKEY EGG-WHITE LYSOZYME, 2 ANTIBODY-PROTEIN COMPLEX, SINGLE-CHAIN FV FRAGMENT
1667	1l3r	B	46	285	3.2e-68	0.49	-0.19		ACETYLCHOLINE RECEPTOR	IMMUNE SYSTEM IG-FOLD,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Pi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1667	1lr	L	46	286	3.2e-54			51.40	ALPHA: CHAIN: A; FV ANTIBODY FRAGMENT; CHAIN: B;	IMMUNO COMPLEX, ANTIBODY-ANTIGEN, BETA-TURN
1667	1fns	L	46	220	4.8e-56	0.24	-0.03		4-4-20 (IG*G2A=KAPPA=) FAB FRAGMENT; IFLR 5 CHAIN: L, H; IFLR 6	IMMUNOGLOBULIN
1667	1fvd	A	46	205	1.6e-56	0.15	-0.08		IMMUNOGLOBULIN NMC-4 IGG1; CHAIN: L; IMMUNOGLOBULIN NMC-4 IGG1; CHAIN: H; VON WILLEBRAND FACTOR; CHAIN: A;	IMMUNE SYSTEM VON WILLEBRAND FACTOR, GLYCOPROTEIN BA (A:ALPHA) BINDING, 2 COMPLEX (WILLEBRAND/IMMUNOGLOBULIN), BLOOD COAGULATION TYPE 3 2B VON WILLEBRAND DISEASE
1667	1gat	H	164	359	1.6e-44	0.00	-0.20		IMMUNOGLOBULIN FAB FRAGMENT OF HUMANIZED ANTIBODY 4D5, VERSION 4 IFVD 3	
1667	1ghf	L	46	283	4.8e-52			51.45	CHIMERIC 48G7 FAB; CHAIN: H, L;	CATALYTIC ANTIBODY ESTER HYDROLYSIS, ESTEROLYTIC, FAB, CATALYTIC ANTIBODY
1667	1epo	L	46	283	8e-52			50.09	ANTI-ANTI-IDIO TYPE GH1002 FAB FRAGMENT; CHAIN: L, H M, I;	ANTIBODY FAB FRAGMENT ANTIBODY FAB FRAGMENT IMMUNOGLOBULIN PROTEIN ENGINEERING, ANTIBODY DESIGN, IMMUNOGLOBULIN 2 STRUCTURE, ANTIGEN-BINDING SITE, CANONICAL

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1667	1h1l	A	46	220	6.4e-60	0.16	-0.11		IMMUNOGLOBULIN IGG2A FAB FRAGMENT (FAB 17/9) 1H1L 3	CONFORMATION, 3 COMPLEMENTARITY- DETERMINING REGION
1667	1h1l	A	46	283	6.4e-60			51.93	IMMUNOGLOBULIN IGG2A FAB FRAGMENT (FAB 17/9) 1H1L 3	
1667	1ifh	L	46	220	6.4e-60	0.22	-0.12		IMMUNOGLOBULIN IGG2A FAB FRAGMENT (FAB 17/9) COMPLEX WITH PEPTIDE OF 1IFH 3 INFLUENZA HEMAGGLUTININ HA1 (STRAIN X47) (RESIDUES 101-107) 1IFH 4	
1667	1ifh	L	46	283	6.4e-60			52.40	IMMUNOGLOBULIN IGG2A FAB FRAGMENT (FAB 17/9) COMPLEX WITH PEPTIDE OF 1IFH 3 INFLUENZA HEMAGGLUTININ HA1 (STRAIN X47) (RESIDUES 101-107) 1IFH 4	
1667	1igc	L	46	286	1.6e-52			50.01	COMPLEX (ANTIBODY/BINDING PROTEIN) IGG1 FAB FRAGMENT COMPLEXED WITH PROTEIN G (DOMAIN	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1667	1igt	A	46	286	9.6e-57			50.70	III) IIGC 5 PROTEIN G, STREPTOCOCCUS IIGC 15 IGG2A INTACT ANTIBODY - MAB231; CHAIN: A, B, C, D	IMMUNOGLOBULIN INTACT IMMUNOGLOBULIN V REGION C REGION, IMMUNOGLOBULIN
1667	1lmk	A	47	285	1.1e-76	0.34	-0.15		IMMUNOGLOBULIN ANTI-PHOSPHATIDYLINOSITOL SPECIFIC PHOSPHOLIPASE C DIABODY 1LMK 3 SYNONYMS: LSMK16 DIABODY, SINGLE-CHAIN FV DIMER 1LMK 4	
1667	1mcp	L	46	220	9.6e-62	0.28	-0.13		IMMUNOGLOBULIN IMMUNOGLOBULIN FAB FRAGMENT (MC/PC\$603) 1MCP 4	
1667	1mcp	L	46	286	9.6e-62			55.95	IMMUNOGLOBULIN IMMUNOGLOBULIN FAB FRAGMENT (MC/PC\$603) 1MCP 4	
1667	1nca	L	46	220	1.6e-56	0.09	-0.14		HYDROLASE(O-GLYCOSYL) N9 NEURAMINIDASE-NC41 (E.C.3.2.1.18) COMPLEX WITH FAB 1NCA 3	
1667	1nqb	A	47	286	4.8e-87	0.21	-0.18		SINGLE-CHAIN ANTIBODY FRAGMENT; CHAIN: A, C;	IMMUNOGLOBULIN VARIABLE HEAVY (VH) DOMAIN, VARIABLE LIGHT (VL) ANTIBODY FRAGMENT,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1667	1nsn	L	46	220	4.8e-57	0.21	-0.17		IGG FAB (IGG1, KAPPA); INSN 4 CHAIN: L, H; INSN 5 STAPHYLOCOCCAL NUCLEASE; INSN 9 CHAIN: S; INSN 10	MULTIVALENT ANTIBODY, DIABODY, DOMAIN 2 SWAPPING, IMMUNOGLOBULIN COMPLEX (IMMUNOGLOBULIN/HYDROLASE) N10 FAB IMMUNOGLOBULIN: INSN 7 STAPHYLOCOCCAL NUCLEASE RIBONUCLEASE, INSN 11 IMMUNOGLOBULIN, STAPHYLOCOCCAL NUCLEASE INSN 25
1667	1plg	L	46	282	6.4e-53			50.29	IGG2A=KAPPA=; 1PLG 4 CHAIN: L, H; 1PLG 5	IMMUNOGLOBULIN
1667	1qok	A	46	285	1.6e-80	0.43	-0.13		MFE-23 RECOMBINANT ANTIBODY FRAGMENT; CHAIN: A;	IMMUNOGLOBULIN IMMUNOGLOBULIN, SINGLE-CHAIN FV, ANTI-CARCINOEMBRYONIC 2 ANTIGEN
1667	1sbs	L	46	220	8e-63	0.13	-0.14		MONOCLONAL ANTIBODY 3A2; CHAIN: H, L;	MONOCLONAL ANTIBODY MONOCLONAL ANTIBODY, FAB-FRAGMENT, REPRODUCTION
1667	2fgw	L	46	205	9.6e-58	0.42	-0.13		IMMUNOGLOBULIN FAB FRAGMENT OF A HUMANIZED VERSION OF THE ANTI-CD18 2FGW 3 ANTIBODY H52' (HUH52-OZ	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Pst Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1667	32c2	A	46	220	1.3e-56	0.13	-0.07		FAB/2FGW 4 IGG1 ANTIBODY 32C2; CHAIN: A; IGG1 ANTIBODY 32C2; CHAIN: B;	IMMUNE SYSTEM FAB, ANTIBODY, AROMATASE, P450
1682	1c0t	A	32	322	1.4e-67	-0.20	0.41		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1682	1c1c	B	32	322	1.3e-84	-0.17	0.63		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1682	1c9r	A	32	322	4.8e-75	0.10	0.99		HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5'-CHAIN: T; DNA (5'-CHAIN: P;	TRANSFERASE/IMMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184LE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184L, TRANSFERASE/IMMUNE 3 SYSTEM/DNA
1682	1c9r	B	12	416	1.3e-82			106.36	HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B);	TRANSFERASE/IMMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184LE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1682	1c9r	B	32	322	1.3e-82	-0.24	0.82		CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5'- CHAIN: T; DNA (5'- CHAIN: P;	3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/MMUNE 3 SYSTEM/DNA
1682	1har		12	219	1.3e-55			71.14	REVERSE TRANSCRIPTASE HIV-1 REVERSE TRANSCRIPTASE (AMINO-TERMINAL HALF) (FINGERS IHAR 3 AND PALM SUBDOMAINS) (RT216) (E.C.2.7.7.49) IHAR 4	TRANSFERASE/MMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184ILE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/MMUNE 3 SYSTEM/DNA
1682	1har		32	219	1.3e-55	0.28	0.99		REVERSE TRANSCRIPTASE HIV-1 REVERSE TRANSCRIPTASE (AMINO-TERMINAL HALF) (FINGERS IHAR 3 AND PALM SUBDOMAINS) (RT216) (E.C.2.7.7.49) IHAR 4	



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1682	1mmrl		1	239	1.4e-54			190.23	MMLV REVERSE TRANSCRIPTASE; IMMML 4 CHAIN: NULL; IMMML 5	REVERSE TRANSCRIPTASE
1682	1mmml		32	238	1.4e-54	0.35	1.00		MMLV REVERSE TRANSCRIPTASE; IMMML 4 CHAIN: NULL; IMMML 5	REVERSE TRANSCRIPTASE
1682	1rth	A	12	416	4.8e-92			63.45	HIV-1 REVERSE TRANSCRIPTASE; 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1682	1rth	A	32	322	4.8e-92	-0.04	0.99		HIV-1 REVERSE TRANSCRIPTASE; 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1682	1rth	B	16	405	3.2e-87			108.51	HIV-1 REVERSE TRANSCRIPTASE; 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1682	1rth	B	32	322	3.2e-87	-0.06	0.65		HIV-1 REVERSE TRANSCRIPTASE; 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1682	1vrt	A	16	416	4.8e-92			69.69	HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
1682	1vrt	A	32	322	4.8e-92	-0.05	1.00		HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
									CHAIN: A, B; 1VRT 5	REVERSE TRANSCRIPTASE 1VRT 15
1682	1vrt	B	16	395	3.2e-86			107.54	HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
1682	1vrt	B	32	322	3.2e-86	-0.02	0.89		HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
1682	3hvt	B	14	395	4.8e-85			110.25	NUCLEOTIDYLTRANSFERASE REVERSE TRANSCRIPTASE (E.C.2.7.49) 3HVT 3	
1682	3hvt	B	32	322	4.8e-85	-0.18	0.90		NUCLEOTIDYLTRANSFERASE REVERSE TRANSCRIPTASE (E.C.2.7.49) 3HVT 3	
1683	1qj	B	23	197	1.4e-19	0.43	0.55		HIS TAG; CHAIN: A; HTLV-1 CAPSID PROTEIN; CHAIN: B;	VIRUS/VIRAL PROTEIN HTLV-1, CAPSID PROTEIN, RETROVIRUS, TWO-DOMAIN PROTEIN, 2 ALPHA HELICAL PROTEIN, HETERONUCLEAR NMR SPECTROSCOPY, 3 VIRUS/VIRAL PROTEIN
1683	1qj	B	34	172	0.00016	-0.05	0.65		HIS TAG; CHAIN: A; HTLV-1 CAPSID PROTEIN; CHAIN: B;	VIRUS/VIRAL PROTEIN HTLV-1, CAPSID PROTEIN, RETROVIRUS, TWO-DOMAIN PROTEIN, 2 ALPHA HELICAL PROTEIN,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1692	1b0	A	43	74	0.0045	-0.60	0.31		UBIQUITIN-LIKE PROTEIN 7, RUB1; CHAIN: A;	SIGNALING PROTEIN RUB1, UBIQUITIN-LIKE PROTEIN, ARABIDOPSIS, SIGNALING PROTEIN
1692	1c3t	A	30	90	4.5e-07	-0.20	0.45		ID8 UBIQUITIN; CHAIN: A;	DE NOVO PROTEIN PROTEIN DESIGN, HYDROPHOBIC CORE, PACKING, ROTAMERS, ROC, 2 UBIQUITIN, DE NOVO PROTEIN, UBIQUITIN
1692	1tbe	B	30	86	4.5e-07	-0.23	0.95		UBIQUITIN TETRAUBIQUITIN 1TBE 3	
1692	1ubi		30	90	9e-08	-0.15	0.46		CHROMOSOMAL PROTEIN UBIQUITIN 1UBI 3	
1692	1ud7	A	30	90	9e-08	-0.12	0.55		UBIQUITIN CORE MUTANT ID7; CHAIN: A;	UBIQUITIN UBIQUITIN, DESIGNED CORE MUTANT
1709	1deq	B	38	253	1.8e-84	0.38	0.70		FIBRINOGEN (ALPHA CHAIN); CHAIN: A, D, N, Q; FIBRINOGEN (BETA CHAIN); CHAIN: B, E, O, R; FIBRINOGEN (GAMMA CHAIN); CHAIN: C, F, P, S; FIBRINOGEN; CHAIN: M, Z;	BLOOD CLOTTING COILED-COIL
1709	1deq	B	40	254	3.2e-82	0.56	0.71		FIBRINOGEN (ALPHA CHAIN);	BLOOD CLOTTING COILED-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN: A, D, N, Q; FIBRINOGEN (BETA CHAIN); CHAIN: B, E, O, R; FIBRINOGEN (GAMMA CHAIN); CHAIN: C, F, P, S; FIBRINOGEN; CHAIN: M, Z;	COIL
1709	1deq	C	39	254	1.6e-89	0.50	0.94		FIBRINOGEN (ALPHA CHAIN); CHAIN: A, D, N, Q; FIBRINOGEN (BETA CHAIN); CHAIN: B, E, O, R; FIBRINOGEN (GAMMA CHAIN); CHAIN: C, F, P, S; FIBRINOGEN; CHAIN: M, Z;	BLOOD CLOTTING COILED-COIL
1709	1ei3	B	39	254	1.6e-86	0.30	0.81		FIBRINOGEN; CHAIN: A, D; FIBRINOGEN; CHAIN: B, E; FIBRINOGEN; CHAIN: C, F;	BLOOD CLOTTING COILED COILS, DISULFIDE RINGS, FIBRIN FORMING ENTITIES
1709	1ei3	C	39	254	1.4e-89	0.58	1.00		FIBRINOGEN; CHAIN: A, D; FIBRINOGEN; CHAIN: B, E; FIBRINOGEN; CHAIN: C, F;	BLOOD CLOTTING COILED COILS, DISULFIDE RINGS, FIBRIN FORMING ENTITIES
1709	1ffb		36	252	4.8e-90			173.33	GAMMA-FIBRINOGEN CARBOXYL TERMINAL FRAGMENT; CHAIN: NULL;	BLOOD COAGULATION FACTOR BLOOD COAGULATION, GLYCOPROTEIN, CALCIUM, PLATELET, PLASMA, 2
1709	1ffb		40	254	4.8e-90	0.57	1.00		GAMMA-FIBRINOGEN	SIGNAL, DISEASE MUTATION, 3 POLYMORPHISM BLOOD COAGULATION

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
									CARBOXYL TERMINAL FRAGMENT; CHAIN: NULL;	FACTOR BLOOD COAGULATION; GLYCOPROTEIN, CALCIUM, PLATELET, PLASMA, 2 ALTERNATIVE SPLICING, SIGNAL, DISEASE MUTATION, 3 POLYMORPHISM
1709	1fzc	B	1	254	1.1e-86			150.56	FIBRIN; CHAIN: A, B, C, D, E, F, G, H, I, J;	BLOOD COAGULATION BLOOD COAGULATION, PLASMA PROTEIN, CROSSLINKING
1709	1fzc	B	40	254	1.1e-86	0.75	1.00		FIBRIN; CHAIN: A, B, C, D, E, F, G, H, I, J;	BLOOD COAGULATION BLOOD COAGULATION, PLASMA PROTEIN, CROSSLINKING
1709	1fzc	C	1	252	4.8e-90			163.73	FIBRIN; CHAIN: A, B, C, D, E, F, G, H, I, J;	BLOOD COAGULATION BLOOD COAGULATION, PLASMA PROTEIN, CROSSLINKING
1709	1fzc	C	40	254	4.8e-90	0.80	1.00		FIBRIN; CHAIN: A, B, C, D, E, F, G, H, I, J;	BLOOD COAGULATION BLOOD COAGULATION, PLASMA PROTEIN, CROSSLINKING
1709	1fzd	A	77	255	6.4e-76			135.85	FIBRINOGEN-420; CHAIN: A, B, C, D, E, F, G, H;	BLOOD COAGULATION BLOOD COAGULATION, FIBRINOGEN-420, ALPHA3C DOMAIN, 2 FIBRINOGEN RELATED DOMAIN, GLYCOSYLATED PROTEIN
1709	1fzg	C	1	252	4.8e-90			169.27	FIBRINOGEN; CHAIN: A, B, C, D, E, F, S, T, M, N;	BLOOD COAGULATION BLOOD COAGULATION, PLASMA, PLATELET, FIBRINOGEN,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1709	1fzg	C	40	254	4.8e-90	0.72	1.00		FIBRINOGEN; CHAIN: A, B, C, D, E, F, S, T, M, N;	FIBRIN BLOOD COAGULATION BLOOD COAGULATION, PLASMA, PLATELET, FIBRINOGEN, FIBRIN
1709	1fzg	E	1	253	1.1e-86			155.95	FIBRINOGEN; CHAIN: A, B, C, D, E, F, S, T, M, N;	BLOOD COAGULATION BLOOD COAGULATION, PLASMA, PLATELET, FIBRINOGEN, FIBRIN
1709	1fzg	E	40	254	1.1e-86	0.67	1.00		FIBRINOGEN; CHAIN: A, B, C, D, E, F, S, T, M, N;	BLOOD COAGULATION BLOOD COAGULATION, PLASMA, PLATELET, FIBRINOGEN, FIBRIN
1712	1a06		2	57	3.2e-18	-0.33	0.18		CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE; CHAIN: NULL;	KINASE KINASE, SIGNAL TRANSDUCTION, CALCIUM/CALMODULIN
1712	1apm	E	2	58	1.4e-20	-0.37	0.48		TRANSFERRASE/PHOSPHOTRANSFERRASE) \$C./AMP\$-DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (\$C/APPK\$) 1APM 3 (CATALYTIC SUBUNIT) "ALPHA" ISOENZYME MUTANT WITH SER 139 1APM 4 REPLACED BY ALA (/S139A\$) COMPLEX WITH THE PEPTIDE 1APM 5	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									INHIBITOR PKI(5-24) AND THE DETERGENT MEGA-8 1APM 6	
1712	1cmk	E	2	58	1.4e-20	-0.31	0.70		PHOSPHOTRANSFERASE CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT 1CMK 3 (E.C.2.7.1.37) 1CMK 4	
1712	1ctp	B	2	58	1.4e-20	-0.45	0.57		TRANSFERASE(PHOSPHOTRANSFERASE) CAMP-DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (CAPK) 1CTP 3 (CATALYTIC SUBUNIT) 1CTP 4	
1712	1koa		1	57	8e-13	-0.31	0.28		TWITCHIN; CHAIN: NULL;	KINASE KINASE, TWITCHIN, INTRASTERIC REGULATION
1712	1kob	A	1	57	3.2e-12	-0.71	0.11		TWITCHIN; CHAIN: A, B;	KINASE KINASE, TWITCHIN, INTRASTERIC REGULATION
1712	1phk		1	54	3.2e-14	-0.36	0.45		PHOSPHORYLASE KINASE; CHAIN: NULL;	KINASE RABBIT MUSCLE PHOSPHORYLASE KINASE; GLYCOGEN METABOLISM, TRANSFERASE, SERINE/THREONINE-PROTEIN, 2 KINASE, ATP-BINDING, CALMODULIN-BINDING
1712	1kd	A	1	55	9.6e-12	-0.74	0.33		TTTN; CHAIN: A, B;	SERINE KINASE SERINE KINASE, TITIN, MUSCLE, AUTOINHIBITION
1715	1ez3	A	101	155	2.2e-08	1.09	-0.19		SYNTAXIN-1A; CHAIN: A, B;	ENDOCYTOSIS/EXOCYTOSIS

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1715	1ez3	A	98	156	4.5e-11	0.89	-0.20		C; SYNTAXIN-1A; CHAIN: A, B, C;	SYNAPTOTAGMIN ASSOCIATED, 35 KDA PROTEIN, P35A, THREE HELIX BUNDLE ENDOCYTOSIS/EXOCYTOSIS SYNAPTOTAGMIN ASSOCIATED 35 KDA PROTEIN, P35A, THREE HELIX BUNDLE
1715	2tc	P	100	164	2.7e-09	0.24	-0.20		TRANSDUCIN; CHAIN: B, G; PHOSDUCIN; CHAIN: P;	COMPLEX (TRANSDUCER/TRANSDUCTION ) GT BETA-GAMMA; MEKA, PP33; PHOSDUCIN, TRANSDUCIN, BETA-GAMMA, SIGNAL TRANSDUCTION, 2 REGULATION, PHOSPHORYLATION, G PROTEINS, THIOREDOXIN, 3 VISION, MEKA, COMPLEX (TRANSDUCER/TRANSDUCTION )
1715	2tc	P	102	156	3.2e-12	0.32	-0.19		TRANSDUCIN; CHAIN: B, G; PHOSDUCIN; CHAIN: P;	COMPLEX (TRANSDUCER/TRANSDUCTION ) GT BETA-GAMMA; MEKA, PP33; PHOSDUCIN, TRANSDUCIN, BETA-GAMMA, SIGNAL TRANSDUCTION, 2 REGULATION, PHOSPHORYLATION, G PROTEINS, THIOREDOXIN, 3



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1716	1bm4	A	166	196	1.1e-06	-0.60	0.13		MOONEY MURINE LEUKEMIA VIRUS CAPSID; CHAIN: A;	VIRUS/VIRAL PROTEIN MOMLV CA MHR PEPTIDE ANALOG; MOONEY MURINE LEUKEMIA VIRUS CAPSID PROTEIN, MOMLV, MU-MLV, 2 CAPSID, MHR, MAJOR HOMOLOG REGION, VIRUS/VIRAL PROTEIN
1716	1d1d	A	52	225	8e-23	0.10	-0.06		CAPSID PROTEIN; CHAIN: A;	VIRUS/VIRAL PROTEIN TWO INDEPENDENT DOMAINS HELICAL BUNDLES, VIRUS/VIRAL PROTEIN
1716	1e0q	A	169	224	4.8e-08	0.05	-0.06		GAG POLYPEPTIDE CAPSID PROTEIN P27; CHAIN: A;	VIRUS/VIRAL PROTEIN VIRUS/VIRAL PROTEIN
1722	1dun		96	203	1.1e-25	0.14	0.77		DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE; CHAIN: NULL;	HYDROLASE DUTPASE, DUTP PYROPHOSPHATASE; HYDROLASE, DUTPASE, ELAV, TRIMERIC ENZYME, ASPARTYL PROTEASE
1722	1euw	A	84	204	4.8e-27	0.13	0.15		DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE; CHAIN: A;	HYDROLASE DUTPASE; JELLY ROLL, MERCURY DERIVATIVE
1722	1f7d	A	101	200	1.4e-27	0.07	0.89		POL. POLYPEPTIDE; CHAIN: A;	VIRUS/VIRAL PROTEIN EIGHT

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1722	1f7r	A	101	220	4.8e-35	-0.12	0.25		B; POL POLYPEPTIDE; CHAIN: A;	STRANDED BETA-BARREL VIRUS/VIRAL PROTEIN EIGHT STRANDED BETA BARREL PROTEIN
1725	1b1h	A	32	146	4.8e-20	0.08	-0.12		HEMOLIN; CHAIN: A, B;	INSECT IMMUNITY INSECT IMMUNITY, LPS-BINDING, HOMOPHILIC ADHESION
1725	1b1h	A	65	436	1.1e-22			92.63	HEMOLIN; CHAIN: A, B;	INSECT IMMUNITY INSECT IMMUNITY, LPS-BINDING, HOMOPHILIC ADHESION
1725	1b12	A	211	321	0.0009	-0.39	0.31		HLA-DR2; CHAIN: A, D; HLA- DR2; CHAIN: B, E; HLA-DR2; CHAIN: C, F;	IMMUNE SYSTEM HLA-DR2, MYELIN BASIC PROTEIN, MULTIPLE SCLEROSIS, 2 AUTOIMMUNITY, IMMUNE SYSTEM
1725	1cvs	C	189	335	9.6e-24	0.05	0.47		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH FACTOR/GROWTH FACTOR RECEPTOR
1725	1cvs	C	57	150	1.6e-21	0.07	0.54		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1725	1evs	D	189	335	9.6e-23	-0.07	0.40		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	FACTOR/GROWTH FACTOR RECEPTOR
1725	1epf	A	30	140	1.4e-16	-0.24	0.17		NEURAL CELL ADHESION MOLECULE; CHAIN: A, B, C, D;	CELL ADHESION NCAM; NCAM, IMMUNOGLOBULIN FOLD, GLYCOPROTEIN
1725	1ev2	E	189	335	1.4e-18	-0.21	0.19		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF2; FGF2; IMMUNOGLOBULIN (IG)-LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
1725	1ev2	E	242	335	6.4e-23	0.32	0.75		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF2; FGF2; IMMUNOGLOBULIN (IG)-LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
1725	1ev2	B	56	242	6.4e-25	-0.29	0.03		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF2;

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	FGFR2; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
1725	lev2	G	189	337	1.6e-19	0.00	0.47		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGFR2; FGFR2; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
1725	lev2	G	242	335	6.4e-23	0.19	0.62		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGFR2; FGFR2; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
1725	lev1	C	189	335	4.8e-22	0.01	0.40		FIBROBLAST GROWTH FACTOR 1; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGFR1; FGFR1; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
1725	lev1	C	240	335	1.8e-22	0.19	0.55		FIBROBLAST GROWTH FACTOR 1; CHAIN: A, B;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGFR1;

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	FGFR1; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREEFOLD
1725	1evt	C	243	335	3.2e-22	-0.09	0.69		FIBROBLAST GROWTH FACTOR 1; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGFR1; FGFR1; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREEFOLD
1725	1fhg	A	237	335	4.5e-28	0.52	1.00		TELOKIN; CHAIN: A	CONTRACTILE PROTEIN IMMUNOGLOBULIN FOLD, BETA BARREL
1725	1fhg	A	237	338	6.4e-28	0.51	0.96		TELOKIN; CHAIN: A	CONTRACTILE PROTEIN IMMUNOGLOBULIN FOLD, BETA BARREL
1725	1fhg	A	64	150	1.6e-24	-0.04	0.21		TELOKIN; CHAIN: A	CONTRACTILE PROTEIN IMMUNOGLOBULIN FOLD, BETA BARREL
1725	1fv1	A	211	321	0.0009	-0.60	0.18		MAJOR HISTOCOMPATIBILITY COMPLEX ALPHA CHAIN; CHAIN: A, D; MAJOR HISTOCOMPATIBILITY COMPLEX BETA CHAIN; CHAIN: B, E; MYELIN BASIC	IMMUNE SYSTEM MHC CLASS II DR2A

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1725	1hdm	B	195	323	0.0013	-0.29	0.36		PROTEIN; CHAIN: C, F; CLASS II HISTOCOMPATIBILITY ANTIGEN, M ALPHA CHAIN: A; CLASS II HISTOCOMPATIBILITY ANTIGEN, M BETA CHAIN: B;	IMMUNE SYSTEM RING6, HLA- DMA; RING7, HLA-DMB; HISTOCOMPATIBILITY PROTEIN, IMMUNE SYSTEM
1725	1ib	B	68	338	3.1e-23			57.87	INTERLEUKIN-1 BETA; CHAIN: A; TYPE 1 INTERLEUKIN-1 RECEPTOR; CHAIN: B;	COMPLEX (IMMUNOGLOBULIN/RECEPTOR ) (IMMUNOGLOBULIN FOLD, TRANSMEMBRANE, GLYCOPROTEIN, RECEPTOR, 2 SIGNAL, COMPLEX (IMMUNOGLOBULIN/RECEPTOR )
1725	1nct		239	336	2.7e-26			54.97	TTTN; CHAIN: NULL;	MUSCLE PROTEIN CONNECTIN, NEXTMS; CELL ADHESION, GLYCOPROTEIN, TRANSMEMBRANE, REPEAT, BRAIN, 2 IMMUNOGLOBULIN FOLD, ALTERNATIVE SPLICING, SIGNAL, 3 MUSCLE PROTEIN
1725	1nct		241	335	2.7e-26	0.04	0.96		TTTN; CHAIN: NULL;	MUSCLE PROTEIN CONNECTIN, NEXTMS; CELL ADHESION, GLYCOPROTEIN, TRANSMEMBRANE, REPEAT, BRAIN, 2 IMMUNOGLOBULIN FOLD, ALTERNATIVE SPLICING,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1725	2fc6	A	250	365	8e-09	-0.44	0.19		FC GAMMA RIIB; CHAIN: A;	SIGNAL, 3 MUSCLE PROTEIN IMMUNE SYSTEM CD32; RECEPTOR, FC, CD32, IMMUNE SYSTEM
1725	3ncm	A	245	335	1.8e-24	0.29	0.17		NEURAL CELL ADHESION MOLECULE, LARGE ISOFORM; CHAIN: A;	CELL ADHESION PROTEIN NCAM MODULE 2; CELL ADHESION, GLYCOPROTEIN, HEPARIN-BINDING, GPI-ANCHOR, 2 NEURAL ADHESION MOLECULE, IMMUNOGLOBULIN FOLD, HOMOPHILIC 3 BINDING, CELL ADHESION PROTEIN
1727	1a06		62	266	6.4e-55	-0.05	0.96		CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE; CHAIN: NULL;	KINASE KINASE, SIGNAL TRANSDUCTION, CALCIUM/CALMODULIN
1727	1apm	B	41	267	1.4e-83			74.69	TRANSFERASE(PHOSPHOTRANSFERASE) \$C-/AMP\$-DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (\$C/APKS) 1APM 3 (CATALYTIC SUBUNIT) "ALPHA" ISOENZYME MUTANT WITH SER 139 1APM 4 REPLACED BY ALA (/S139AS) COMPLEX WITH THE PEPTIDE 1APM 5	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1727	1apm	E	63	265	1.4e-83	0.55	1.00		INHIBITOR PKI(5-24) AND THE DETERGENT MEGA-8 1APM 6	
									TRANSFERASE/PHOSPHOTRANSFERASE) \$C-/AMPS-DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (\$CAPK\$) 1APM 3 (CATALYTIC SUBUNIT) "ALPHA" ISOENZYME MUTANT WITH SER 139 1APM 4 REPLACED BY ALA (/S139A\$) COMPLEX WITH THE PEPTIDE 1APM 5 INHIBITOR PKI(5-24) AND THE DETERGENT MEGA-8 1APM 6	
1727	1cmk	E	32	267	1.4e-85			76.30	PHOSPHOTRANSFERASE CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT 1CMK 3 (E.C.2.7.1.37) 1CMK 4	
1727	1cmk	E	63	265	1.4e-85	0.50	1.00		PHOSPHOTRANSFERASE CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT 1CMK 3 (E.C.2.7.1.37) 1CMK 4	
1727	1cjp	E	39	267	1.4e-85			81.96	TRANSFERASE/PHOSPHOTRANSFERASE) CAMP-DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (CAPK) 1CTP 3 (CATALYTIC	



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1727	1cjp	E	63	265	1.4e-85	0.46	1.00		SUBUNIT1 ICTP 4	
									TRANSFERASE(PHOSPHOTRANSFERASE) CAMP-DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (CAPK) ICTP 3 (CATALYTIC SUBUNIT) ICTP 4	
1727	1f3m	C	68	263	3.1e-51	0.16	1.00		SERINE/THREONINE-PROTEIN KINASE PAK-ALPHA; CHAIN: A, B; SERINE/THREONINE-PROTEIN KINASE PAK-ALPHA; CHAIN: C, D; TWITCHIN; CHAIN: NULL;	TRANSFERASE KINASE DOMAIN, AUTONHIBITORY FRAGMENT, HOMODIMER
1727	1koa		69	263	2.3e-52	0.31	0.99		TWITCHIN; CHAIN: NULL;	KINASE KINASE, TWITCHIN, INTRASTERIC REGULATION
1727	1kob	A	69	262	2.7e-51	0.03	1.00		TWITCHIN; CHAIN: A, B;	KINASE KINASE, TWITCHIN, INTRASTERIC REGULATION
1727	1phk		68	267	4.5e-62			66.85	PHOSPHORYLASE KINASE; CHAIN: NULL;	KINASE RABBIT MUSCLE PHOSPHORYLASE KINASE; GLYCOGEN METABOLISM, TRANSFERASE, SERINE/THREONINE-PROTEIN, 2 KINASE, ATP-BINDING, CALMODULIN-BINDING
1727	1phk		69	263	4.5e-62	0.39	1.00		PHOSPHORYLASE KINASE; CHAIN: NULL;	KINASE RABBIT MUSCLE PHOSPHORYLASE KINASE; GLYCOGEN METABOLISM, TRANSFERASE, SERINE/THREONINE-PROTEIN, 2

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1727	1phk		69	265	1.3e-56	0.46	1.00		PHOSPHORYLASE KINASE; CHAIN: NULL;	KINASE, ATP-BINDING, CALMODULIN-BINDING
1727	1tki	A	69	263	9e-54	0.08	0.90		TITIN; CHAIN: A, B;	SERINE KINASE SERINE KINASE, TITIN, MUSCLE, AUTOINHIBITION
1732	1d5r	A	48	343	0			143.72	PHOSPHOINOSITIDE PHOSPHOTASE PTEN; CHAIN: A;	HYDROLASE C2 DOMAIN, PHOSPHOTIDYLINOSITOL, PHOSPHOTASE, HYDROLASE
1732	1d5r	A	49	343	0	0.55	1.00		PHOSPHOINOSITIDE PHOSPHOTASE PTEN; CHAIN: A;	HYDROLASE C2 DOMAIN, PHOSPHOTIDYLINOSITOL, PHOSPHOTASE, HYDROLASE
1734	1ae6	L	86	291	1.4e-05			52.34	ANTIBODY CTM01; CHAIN: L, H;	IMMUNOGLOBULIN, FAB FRAGMENT, HUMANISATION
1734	1clz	L	84	291	0.00032			51.55	IGG FAB (GG3, KAPPA); CHAIN: L, H;	IMMUNOGLOBULIN MBR96 FAB (IMMUNOGLOBULIN); IMMUNOGLOBULIN C REGION, GLYCOPROTEIN, TRANSMEMBRANE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1734	1cs6	A	23	310	9.6e-28	-0.18	0.22		AXONIN-1; CHAIN: A;	CELL ADHESION NEURAL CELL ADHESION
1734	1cd8	A	108	269	8e-07	-0.13	0.40		7C8 FAB FRAGMENT; SHORT CHAIN; CHAIN: A, C; 7C8 FAB FRAGMENT; LONG CHAIN; CHAIN: B, D	IMMUNE SYSTEM ABZYME TRANSITION STATE ANALOG, IMMUNE SYSTEM
1734	1cvs	D	114	285	1.3e-17	0.05	0.00		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH FACTOR/GROWTH FACTOR RECEPTOR
1734	1dbb	L	84	284	0.00048			53.02	IMMUNOGLOBULIN FAB FRAGMENT OF THE DB3 ANTI-STERIOD MONOCLONAL ANTIBODY IDBB 3 (GG1, SUBGROUP 2A, KAPPA 1) COMPLEX WITH PROGESTERONE IDBB 4	
1734	1dqg	A	103	269	3.2e-07	0.25	0.12		ANTI-LYSOZYME ANTIBODY HYHEL-63 (LIGHT CHAIN); CHAIN: A, C; ANTI-LYSOZYME ANTIBODY HYHEL-63 (HEAVY CHAIN); CHAIN: B, D;	IMMUNE SYSTEM ANTI-LYSOZYME ANTIBODY, HYHEL-63, HEN EGG WHITE LYSOZYME
1734	1epf	A	107	269	6.4e-11	-0.00	-0.02		NEURAL CELL ADHESION MOLECULE; CHAIN: A, B, C;	CELL ADHESION NCAM; NCAM, IMMUNOGLOBULIN FOLD,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1734	1epf	A	29	170	1.6e-14	-0.24	0.22		D; NEURAL CELL ADHESION MOLECULE; CHAIN: A, B, C, D;	GLYCOPROTEIN CELL ADHESION NCAM; NCAM, IMMUNOGLOBULIN FOLD, GLYCOPROTEIN
1734	1f2q	A	106	285	1.1e-22	0.21	0.23		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN: A;	IMMUNE SYSTEM FC-EPSILON RI-ALPHA; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IGE-BINDING 2 PROTEIN
1734	1f2q	A	22	195	3.2e-47	0.26	1.00		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN: A;	IMMUNE SYSTEM FC-EPSILON RI-ALPHA; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IGE-BINDING 2 PROTEIN
1734	1f6a	A	106	285	3.2e-22	-0.17	0.13		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN: A; IG EPSILON CHAIN C REGION; CHAIN: B, D;	IMMUNE SYSTEM HIGH AFFINITY IGE-FC RECEPTOR, FC(EPSILON) IGE-FC; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IGE-BINDING 2 PROTEIN, IGE ANTIBODY, IGE-FC
1734	1f6a	A	20	190	8e-47	0.29	0.99		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN: A; IG EPSILON CHAIN C REGION; CHAIN: B, D;	IMMUNE SYSTEM HIGH AFFINITY IGE-FC RECEPTOR, FC(EPSILON) IGE-FC; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IGE-BINDING 2 PROTEIN, IGE ANTIBODY, IGE-FC

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1734	1fcg	A	105	287	8e-27	0.19	-0.09		FC RECEPTOR FC(GAMMA)/IL1A; CHAIN: A;	IMMUNE SYSTEM, MEMBRANE PROTEIN CD32; FC RECEPTOR, IMMUNOGLOBULIN, LEUKOCYTE, CD32
1734	1fcg	A	19	192	3.2e-51	0.30	0.99		FC RECEPTOR FC(GAMMA)/IL1A; CHAIN: A;	IMMUNE SYSTEM, MEMBRANE PROTEIN CD32; FC RECEPTOR, IMMUNOGLOBULIN, LEUKOCYTE, CD32
1734	1flr	L	84	291	6.4e-06			52.22	4-4-20 (IG*G2A=KAPPA=) FAB FRAGMENT; 1FLR 5 CHAIN: L, H; 1FLR 6	IMMUNOGLOBULIN
1734	1flr	A	114	285	1.1e-26	-0.00	0.22		LOW AFFINITY IMMUNOGLOBULIN GAMMA FC REGION CHAIN: A;	IMMUNE SYSTEM RECEPTOR BETA SANDWICH, IMMUNOGLOBULIN-LIKE, RECEPTOR
1734	1flr	A	19	191	9.6e-49	0.34	1.00		LOW AFFINITY IMMUNOGLOBULIN GAMMA FC REGION CHAIN: A;	IMMUNE SYSTEM RECEPTOR BETA SANDWICH, IMMUNOGLOBULIN-LIKE, RECEPTOR
1734	1lgr	L	84	291	0.00016			54.97	IMMUNOGLOBULIN IGG1 FAB FRAGMENT (B1312) IIGF 3	
1734	1lrb	B	1	313	4.8e-10			51.11	INTERLEUKIN-1 BETA; CHAIN: A; TYPE 1 INTERLEUKIN-1 RECEPTOR; CHAIN: B;	COMPLEX (IMMUNOGLOBULIN/RECEPTOR ) IMMUNOGLOBULIN FOLD, TRANSMEMBRANE, GLYCOPROTEIN, RECEPTOR, 2 SIGNAL, COMPLEX

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1734	Inkr		104	287	1.4e-26	-0.12	0.05		P58-CL42 KIR; CHAIN: NULL;	(IMMUNOGLOBULIN/RECEPTOR ) INHIBITORY RECEPTOR KILLER CELL INHIBITORY RECEPTOR; INHIBITORY RECEPTOR, NATURAL KILLER CELLS, IMMUNOLOGICAL 2 RECEPTORS, IMMUNOGLOBULIN FOLD
1734	Inkr		195	301	1.1e-14	0.03	-0.15		P58-CL42 KIR; CHAIN: NULL;	INHIBITORY RECEPTOR KILLER CELL INHIBITORY RECEPTOR; INHIBITORY RECEPTOR, NATURAL KILLER CELLS, IMMUNOLOGICAL 2 RECEPTORS, IMMUNOGLOBULIN FOLD
1734	Inkr		21	191	9.6e-30			55.48	P58-CL42 KIR; CHAIN: NULL;	INHIBITORY RECEPTOR KILLER CELL INHIBITORY RECEPTOR; INHIBITORY RECEPTOR, NATURAL KILLER CELLS, IMMUNOLOGICAL 2 RECEPTORS, IMMUNOGLOBULIN FOLD
1734	Inkr		22	191	9.6e-30	-0.40	0.41		P58-CL42 KIR; CHAIN: NULL;	INHIBITORY RECEPTOR KILLER CELL INHIBITORY RECEPTOR; INHIBITORY RECEPTOR, NATURAL KILLER CELLS, IMMUNOLOGICAL 2

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Pst Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1734	1gok	A	106	267	9.6e-09	0.14	-0.14		MFE-23 RECOMBINANT ANTIBODY FRAGMENT; CHAIN: A;	RECEPTORS, IMMUNOGLOBULIN FOLD
1734	2dli	A	101	287	9.6e-28	-0.14	0.05		MHC CLASS I NK CELL RECEPTOR PRECURSOR; CHAIN: A;	IMMUNE SYSTEM P38 NATURAL KILLER CELL RECEPTOR; KIR, NATURAL KILLER RECEPTOR, INHIBITORY RECEPTOR, 2 IMMUNOGLOBULIN
1734	2dli	A	20	190	3.2e-30	-0.37	0.15		MHC CLASS I NK CELL RECEPTOR PRECURSOR; CHAIN: A;	IMMUNE SYSTEM P38 NATURAL KILLER CELL RECEPTOR; KIR, NATURAL KILLER RECEPTOR, INHIBITORY RECEPTOR, 2 IMMUNOGLOBULIN
1734	2fcb	A	19	193	1.4e-52	0.18	0.98		FC GAMMA R1B; CHAIN: A;	IMMUNE SYSTEM CD32; RECEPTOR, FC, CD32, IMMUNE SYSTEM
1737	1a6q		33	299	4.8e-58	0.29	0.07		PHOSPHATASE 2C; CHAIN: NULL;	HYDROLASE CATALYTIC MECHANISM, METALLOENZYME, PROTEIN PHOSPHATASE 2C, 2 SIGNAL TRANSDUCTUIN, X-RAY

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMTF score	SEQFOLD score	Compound	PDB annotation
1738	1aww		95	162	1.6e-15	-0.17	0.11		BRUTON'S TYROSINE KINASE; CHAIN: NULL;	CRYSTALLOGRAPHY, HYDROLASE
1738	1aze	A	104	157	4.8e-15	-0.25	0.23		GRB2; CHAIN: A; SOS; CHAIN: B;	TRANSFERASE ATK, AMGX1, BPK; TYROSINE KINASE, X-LINKED AGAMMAGLOBULINEMIA, XLA, BTK, SH3 2 DOMAIN, TRANSFERASE
1738	1bul	A	104	160	9.6e-15	0.34	0.13		HEMOPOLYETIC CELL KINASE; CHAIN: A, B, C, D, E, F;	COMPLEX (ADAPTOR PROTEIN/PEPTIDE) ASH, GROWTH FACTOR RECEPTOR-BOUND PROTEIN 2; COMPLEX (ADAPTOR PROTEIN/PEPTIDE), SH3 DOMAIN, 2 GUANINE-NUCLEOTIDE RELEASING FACTOR
1738	1e96	B	3	48	4.8e-05	-0.25	0.16		RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1; CHAIN: A; NEUTROPHIL CYTOSOL FACTOR 2 (NCF-2) CHAIN: B;	TRANSFERASE TYROSINE-PROTEIN KINASE, SIGNAL TRANSDUCTION, 2 SH3
1738	1fyn	A	100	160	1.3e-16	-0.00	0.11		PHOSPHOTRANSFERASE FYN; CHAIN: A; 3BP-2; CHAIN: B;	SIGNALING COMPLEX RAC1; P67PHOX; SIGNALING COMPLEX, GTPASE, NADPH OXIDASE, PROTEIN-PROTEIN 2 COMPLEX, TPR MOTIF TRANSFERASE PROTO-ONCOGENE TYROSINE KINASE;



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										PROTO-ONCOGENE, TRANSFERASE, TYROSINE-PROTEIN KINASE, 2 PHOSPHORYLATION, ATP-BINDING, MYRISTYLATION, SH3 DOMAIN, 3 COMPLEX (PHOSPHOTRANSFERASE/PEPTIDE)
1738	1gbq	A	104	157	3.2e-16	-0.03	0.59		GRB2; CHAIN: A; SOS-1; CHAIN: B;	COMPLEX (SIGNAL TRANSDUCTION/PEPTIDE) COMPLEX (SIGNAL TRANSDUCTION/PEPTIDE), SH3 DOMAIN
1738	1gbr	A	104	163	1.6e-16	-0.30	0.15		SIGNAL TRANSDUCTION PROTEIN GROWTH FACTOR RECEPTOR-BOUND PROTEIN 2 (GRB2, N-TERMINAL 1GBR 3 SH3 DOMAIN) COMPLEXED WITH SOS-A PEPTIDE 1GBR 4 (NMR, 29 STRUCTURES) 1GBR 5	
1738	1gfc		100	160	1.6e-18	0.01	0.70		ADAPTOR PROTEIN CONTAINING SH2 AND SH3 GROWTH FACTOR RECEPTOR-BOUND PROTEIN 2 (GRB2) 1GFC 3 (C-TERMINAL SH3 DOMAIN) (NMR, MINIMIZED MEAN	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1738	1gri	A	4	160	1.6e-30	-0.53	0.01		STRUCTURE) 1GFC 4 GROWTH FACTOR BOUND PROTEIN 2; 1GRI 5 CHAIN: A, B; 1GRI 6	SIGNAL TRANSDUCTION ADAPTOR SH2, SH3 1GRI 14
1738	1hsq		97	163	8e-17	-0.04	0.76		PHOSPHORIC DIESTER HYDROLASE PHOSPHOLIPASE C-GAMMA (SH3 DOMAIN) (E.C.3.1.4.11) IHSQ 3 (NMR, MINIMIZED MEAN STRUCTURE) IHSQ 4	
1738	1pwt		100	160	4.8e-15	0.18	0.16		ALPHA SPECTRIN; CHAIN: NULL;	CIRCULAR PERMUTANT PWT, CIRCULAR PERMUTANT, SH3 DOMAIN, CYTOSKELETON
1738	1sem	A	102	160	3.2e-19	0.14	0.33		SEM-5; ISEM 3 CHAIN: A, B; ISEM 5 10-RESIDUE PROLINE- RICH PEPTIDE FROM MSOS ISEM 8 CHAIN: C, D ISEM 10	SIGNAL TRANSDUCTION PROTEIN SRC-HOMOLOGY 3 (SH3) DOMAIN, PEPTIDE- BINDING PROTEIN, ISEM 18 2 GUANINE NUCLEOTIDE EXCHANGE FACTOR ISEM 19
1738	1shf	A	101	160	1.6e-16	0.63	0.18		PHOSPHOTRANSFERASE FYN PROTO-ONCOGENE TYROSINE KINASE (E.C.2.7.1.112) ISHF 3 (SH3 DOMAIN) ISHF 4	
1738	1yca	B	106	154	1.1e-15	-0.36	0.33		P53; CHAIN: A; 53BP2; CHAIN: B;	COMPLEX (ANTI- ONCOGENE/ANKYRIN REPEATS) P53BP2; ANKYRIN REPEATS, SH3, P53, TUMOR

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1738	4hck		104	161	6.4e-15	0.09	0.16		HEMATOPOIETIC CELL KINASE; CHAIN: NULL;	SUPPRESSOR, MULTIGENE 2 FAMILY, NUCLEAR PROTEIN, PHOSPHORYLATION, DISEASE MUTATION, 3 POLYMORPHISM, COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS)
1739	1a8l		16	72	0.0018	-0.71	0.21		PROTEIN DISULFIDE OXIDOREDUCTASE; CHAIN: NULL;	TRANSFERASE HCK, SH3, PROTEIN TYROSINE KINASE, SIGNAL TRANSDUCTION, 2 TRANSFERASE
1740	1c9r	A	97	264	1.4e-08	0.12	0.27		HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5'- CHAIN: T; DNA (5'- CHAIN: P;	TRANSFERASE/MIMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184LE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/MIMUNE 3 SYSTEM/DNA
1740	1l2l	A	131	225	0.00011	0.21	0.27		RIBONUCLEASE H; CHAIN: A;	HYDROLASE RNASE H, NUCLEASE, RNASE H*, RIBONUCLEASE H, METAL-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Pst Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1740	1f21	A	131	264	1.3e-16	0.37	0.75		RIBONUCLEASE HI, CHAIN: A;	BINDING 2 PROTEIN, PROTEIN FOLDING
1740	1hth	A	130	264	9e-12	0.03	0.13		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H DOMAIN OF /HIV-1\$ REVERSE TRANSCRIPTASE IHRH 3	HYDROLASE RNASE H, NUCLEASE, RNASE H*, RIBONUCLEASE H, METAL-BINDING 2 PROTEIN, PROTEIN FOLDING
1740	1rll		131	264	4.5e-17	0.59	0.82		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H (E.C.3.1.26.4) IRL 3	
1754	1dvp	A	49	94	4.8e-14	-0.20	0.28		HEPATOCTE GROWTH FACTOR-REGULATED TYROSINE CHAIN: A;	TRANSFERASE HRS; HRS, VHS, FYVE, ZINC FINGER, SUPERHELIX
1754	1ptq		47	81	0.004	-0.72	0.07		PROTEIN KINASE C DELTA TYPE, IPTQ 4	PHOSPHOTRANSFERASE
1754	1vfy	A	42	99	3.2e-11	0.05	0.12		PHOSPHATIDYLINOSITOL-3-PHOSPHATE BINDING FYVE CHAIN: A;	TRANSPORT PROTEIN FYVE DOMAIN, ENDOSOME MATURATION, INTRACELLULAR TRAFFICKING, 2 TRANSPORT PROTEIN
1754	1bhd	B	51	105	9.6e-17	-0.33	0.30		RAB-3A; CHAIN: A; RABPHILIN-3A; CHAIN: B;	COMPLEX (GTP-BINDING/EFFECTOR) RAS-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
										RELATED PROTEIN RAB3A; COMPLEX (GTP-BINDING/EFFECTOR), G PROTEIN, EFFECTOR, RABCD, 2 SYNAPTIC EXOCYTOSIS, RAB PROTEIN, RAB3A, RABPHILIN
1760	1b3u	A	7	191	4.8e-13	0.17	-0.06		PROTEIN PHOSPHATASE PP2A; CHAIN: A, B;	SCAFFOLD PROTEIN SCAFFOLD PROTEIN, PP2A, PHOSPHORYLATION, HEAT REPEAT
1760	1ee4	A	4	194	4.8e-31	0.14	0.78		KARYOPHERIN ALPHA; CHAIN: A, B; MYC PROTO-ONCOGENE PROTEIN; CHAIN: C, D, E, F;	TRANSPORT PROTEIN SERINE-RICH RNA POLYMERASE I SUPPRESSOR PROTEIN; ARM REPEAT
1760	1ial	A	4	192	3.2e-31	0.37	0.58		IMPORTIN ALPHA; CHAIN: A;	NUCLEAR IMPORT RECEPTOR KARYOPHERIN ALPHA; NUCLEAR IMPORT RECEPTOR, NUCLEAR LOCALIZATION SIGNAL, 2 ARMADILLO REPEATS, AUTOINHIBITION, INTRASTERIC REGULATION
1760	2bct		11	193	9.6e-23	0.35	0.95		BETA-CATENIN; CHAIN: NULL;	STRUCTURAL PROTEIN ARMADILLO REPEAT, BETA-CATENIN, STRUCTURAL PROTEIN
1760	3bct		5	194	1.1e-21	0.22	0.34		BETA-CATENIN; CHAIN: NULL;	ARMADILLO REPEAT ARMADILLO REPEAT, BETA-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1787	1c28	A	107	240	8e-36	0.32	-0.14		30 KD ADIPOCYTE COMPLEMENT-RELATED PROTEIN CHAIN: A, B, C <sub>2</sub>	SERUM PROTEIN ACRP30 C1Q TNF TRIMER ALL-BETA <sub>2</sub> , SERUM PROTEIN
1787	1c28	B	107	219	1.6e-32	0.41	-0.02		30 KD ADIPOCYTE COMPLEMENT-RELATED PROTEIN CHAIN: A, B, C <sub>2</sub>	SERUM PROTEIN ACRP30 C1Q TNF TRIMER ALL-BETA <sub>2</sub> , SERUM PROTEIN
1787	1c28	C	107	220	6.4e-28	0.46	0.05		30 KD ADIPOCYTE COMPLEMENT-RELATED PROTEIN CHAIN: A, B, C <sub>2</sub>	SERUM PROTEIN ACRP30 C1Q TNF TRIMER ALL-BETA <sub>2</sub> , SERUM PROTEIN
1795	1bj1	L	51	162	1.4e-32	0.33	-0.13		FAB FRAGMENT; CHAIN: L, H, J, K; VASCULAR ENDOTHELIAL GROWTH FACTOR; CHAIN: V, W;	COMPLEX (ANTIBODY/ANTIGEN) FAB-12; VEGF; COMPLEX (ANTIBODY/ANTIGEN), ANGIOGENIC FACTOR
1795	1c1o	H	61	208	3.2e-31	0.02	-0.05		A5B7 MONOCLONAL ANTIBODY; CHAIN: L, H <sub>2</sub>	IMMUNOGLOBULIN, FAB-FRAGMENT
1795	1dee	A	51	162	3.2e-33	0.05	-0.17		IGM RF 2A2; CHAIN: A, C, E; IGM RF 2A2; CHAIN: B, D, F; IMMUNOGLOBULIN G BINDING PROTEIN A; CHAIN: G, H;	IMMUNE SYSTEM/FAB-IBP COMPLEX CRYSTAL STRUCTURE 2.7A RESOLUTION BINDING 2 OUTSIDE THE ANTIGEN COMBINING SITE SUPERANTIGEN FAB VH3 3 SPECIFICITY
1795	1ddb	L	51	162	9.6e-32	0.38	-0.08		IMMUNOGLOBULIN 3D6 FAB	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1795	1dql	L	51	159	1.3e-31	0.30	-0.12		1DFB 3 IGM MEZ IMMUNOGLOBULIN; CHAIN: L; IGM MEZ IMMUNOGLOBULIN; CHAIN: H;	IMMUNE SYSTEM IMMUNOGLOBULIN FOLD, ANTIBODY, IGM, FV
1795	1fvd	A	51	162	3.2e-31	0.35	-0.14		IMMUNOGLOBULIN FAB FRAGMENT OF HUMANIZED ANTIBODY 4D5, VERSION 4 IFVD 3	
1795	1hvx	H	61	208	3.2e-31	0.10	0.04		IMMUNOGLOBULIN 6D9; CHAIN: L, H;	CATALYTIC ANTIBODY 6D9 CATALYTIC ANTIBODY, ESTER HYDROLYSIS, ESTEROLYTIC, FAB, 2 IMMUNOGLOBULIN
1795	2fgw	L	51	162	9.6e-32	0.77	-0.17		IMMUNOGLOBULIN FAB FRAGMENT OF A HUMANIZED VERSION OF THE ANTI-CD18 2FGW 3 ANTIBODY Y52 (HUH52-OZ FAB) 2FGW 4	
1796	1asu		15	129	8e-24	-0.01	0.01		AVIAN SARCOMA VIRUS INTEGRASE; 1ASU 7 CHAIN: NULL, 1ASU 8	DNA INTEGRATION
1796	1asu		21	154	1.7e-26	0.03	0.89		AVIAN SARCOMA VIRUS INTEGRASE; 1ASU 7 CHAIN: NULL, 1ASU 8	DNA INTEGRATION

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1796	1b9d	A	27	129	8e-23	-0.06	0.34		INTEGRASE; CHAIN: A;	TRANSFERASE DNA INTEGRATION
1796	1b9d	A	28	154	6.8e-22	0.09	0.13		INTEGRASE; CHAIN: A;	TRANSFERASE DNA INTEGRATION
1796	1b9f	A	23	129	3.2e-29	-0.51	0.13		INTEGRASE; CHAIN: A;	TRANSFERASE DNA INTEGRATION, TRASFERASE
1796	1b9f	A	24	175	3.4e-26	-0.06	0.40		INTEGRASE; CHAIN: A;	TRANSFERASE DNA INTEGRATION, TRASFERASE
1796	1b13	C	18	129	9.6e-31	-0.30	0.33		INTEGRASE; CHAIN: A, B, C;	DNA INTEGRATION DNA INTEGRATION, AIDS, POLYPROTEIN, HYDROLASE, 2 ENDONUCLEASE, POLYNUCLEOTIDYL TRANSFERASE, DNA BINDING 3 (VIRAL)
1796	1c0m	A	16	129	3.2e-24	0.07	-0.08		INTEGRASE; CHAIN: A, B, C, D;	TRANSFERASE INTEGRASE, ROUS SARCOMA VIRUS, HIV, X-RAY CRYSTALLOGRAPHY, 2 PROTEIN STRUCTURE, TRANSFERASE
1796	1c0m	A	21	175	3.4e-31	0.10	0.49		INTEGRASE; CHAIN: A, B, C, D;	TRANSFERASE INTEGRASE, ROUS SARCOMA VIRUS, HIV, X-RAY CRYSTALLOGRAPHY, 2 PROTEIN STRUCTURE, TRANSFERASE
1796	1c1a	B	24	129	8e-23	-0.01	0.42		RSV INTEGRASE; CHAIN: A, B;	VIRUS/VIRAL PROTEIN INTEGRASE, ROUS SARCOMA VIRUS, HIV, X-RAY



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1796	1c1a	B	24	178	1.4e-25	0.16	0.30		RSV INTEGRASE; CHAIN: A, B;	CRYSTALLOGRAPHY, 2 VIRUS/VIRAL PROTEIN
1796	1cxq	A	21	154	1.7e-26	0.33	0.80		AVIAN SARCOMA VIRUS INTEGRASE; CHAIN: A;	INTEGRASE, ROUS SARCOMA VIRUS, HIV, X-RAY CRYSTALLOGRAPHY, 2 VIRUS/VIRAL PROTEIN
1796	1cz9	A	26	129	1.3e-20	0.14	0.28		AVIAN SARCOMA VIRUS INTEGRASE; CHAIN: A;	TRANSFERASE MIXED BETA-SHEET SURROUNDED BY ALPHA-HELICES
1796	1cz9	A	26	154	1e-26	0.23	0.69		AVIAN SARCOMA VIRUS INTEGRASE; CHAIN: A;	TRANSFERASE MIXED BETA-SHEET SURROUNDED BY ALPHA-HELICES
1796	1cxq	A	27	129	1.6e-22	-0.16	0.16		POLYPROTEIN; CHAIN: A, B;	VIRUS/VIRAL PROTEIN HIV-1 INTEGRASE
1796	1qs4	A	23	129	8e-25	-0.29	0.41		HIV-1 INTEGRASE; CHAIN: A, B, C;	POLYNUCLEOTIDYL TRANSFERASE, DNA-BINDING 2 PROTEIN, DD35E
1796	1qs4	A	24	175	1e-25	-0.19	0.36		HIV-1 INTEGRASE; CHAIN: A, B, C;	HYDROLASE DNA INTEGRATION, INTEGRASE, HIV, HYDROLASE, ASPARTYL 2 PROTEASE, ENDONUCLEASE
1796	1qs4	A	24	175	1e-25	-0.19	0.36		HIV-1 INTEGRASE; CHAIN: A, B, C;	HYDROLASE DNA INTEGRATION, INTEGRASE, HIV, HYDROLASE, ASPARTYL 2

Table 5

SEQ ID NO.	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1802	1mey	C	59	130	1e-25	-0.39	0.34		DNA; CHAIN: A, B, D, E; CONSENSUS ZINC FINGER PROTEIN; CHAIN: C, F, G;	COMPLEX (ZINC FINGER/DNA) ZINC FINGER, PROTEIN-DNA INTERACTION, PROTEIN DESIGN, 2 CRYSTAL STRUCTURE, COMPLEX (ZINC FINGER/DNA)
1802	2gli	A	59	130	6.8e-23	-0.09	0.63		ZINC FINGER PROTEIN GLI1; CHAIN: A; DNA; CHAIN: C, D;	COMPLEX (DNA-BINDING PROTEIN/DNA) FIVE-FINGER GLI, GLI, ZINC FINGER, COMPLEX (DNA-BINDING PROTEIN/DNA)
1806	1kst		29	133	2e-12	0.16	0.41		GELATION FACTOR; CHAIN: NULL;	ACTIN BINDING PROTEIN ABP-120; ACTIN BINDING PROTEIN, STRUCTURE, IMMUNOGLOBULIN, GELATION 2 FACTOR, ABP-120
1806	1qth	A	27	133	2e-11	-0.09	0.53		GELATION FACTOR; CHAIN: A, B;	ACTIN BINDING PROTEIN ACTIN BINDING PROTEIN 120; ACTIN BINDING PROTEIN, IMMUNOGLOBULIN, GELATION FACTOR, ABP-2 120
1812	1ase		153	270	2.4e-21	0.23	0.90		TUMOR SUPPRESSOR P16INK4A; CHAIN: NULL;	ANTI-ONCOGENE CELL CYCLE, ANTI-ONCOGENE, REPEAT, ANK REPEAT

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1812	1awc	B	129	300	1.6e-36	0.63	0.96		GA BINDING PROTEIN ALPHA: CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABP ALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA- BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
1812	1awc	B	164	312	1.4e-29	0.23	0.96		GA BINDING PROTEIN ALPHA: CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABP ALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA- BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
1812	1awc	B	76	225	4.8e-31	0.14	0.45		GA BINDING PROTEIN ALPHA: CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABP ALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA- BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
1812	1awc	B	99	264	1.4e-33	0.38	0.96		GA BINDING PROTEIN	COMPLEX (TRANSCRIPTION

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									ALPHA; CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	REGULATION/DNA) GABPALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA-BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
1812	1bd8		102	267	1.4e-27	0.22	1.00		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
1812	1bd8		132	303	1.6e-26	0.36	0.25		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
1812	1bd8		75	225	3.2e-21	0.02	0.17		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
1812	1blx	B	102	267	1.6e-26	0.38	0.99		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA, COMPLEX (INHIBITOR PROTEIN/KINASE)
1812	1blx	B	132	303	4.8e-27	0.29	0.95		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1812	1bu9	A	102	269	1.4e-27	0.49	0.92		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A;	CONTROL, ALPHA/BETA, COMPLEX (INHIBITOR PROTEIN/KINASE)
1812	1bu9	A	129	305	3.2e-32	0.55	0.58		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A;	HORMONE/GROWTH FACTOR P18-INK4C; CELL CYCLE INHIBITOR, P18INK4C, TUMOR, SUPPRESSOR, CYCLIN-2 DEPENDENT KINASE, HORMONE/GROWTH FACTOR
1812	1bu9	A	167	314	4.8e-25	0.41	0.06		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A;	HORMONE/GROWTH FACTOR P18-INK4C; CELL CYCLE INHIBITOR, P18INK4C, TUMOR, SUPPRESSOR, CYCLIN-2 DEPENDENT KINASE, HORMONE/GROWTH FACTOR
1812	1bu9	A	75	230	1.6e-26	0.12	0.23		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A;	HORMONE/GROWTH FACTOR P18-INK4C; CELL CYCLE INHIBITOR, P18INK4C, TUMOR, SUPPRESSOR, CYCLIN-2 DEPENDENT KINASE, HORMONE/GROWTH FACTOR
1812	1d9s	A	149	271	6.8e-24	0.47	0.95		CYCLIN-DEPENDENT KINASE	SIGNALING PROTEIN HELIX-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1812	1dcq	A	153	276	6.8e-22	0.25	0.69		4 INHIBITOR B; CHAIN: A;	TURN-HELIX, ANKYRIN REPEAT
1812	1ihb	A	102	268	6.4e-27	0.34	1.00		PYK2-ASSOCIATED PROTEIN BETA; CHAIN: A;	METAL BINDING PROTEIN ZINC-BINDING MODULE, ANKYRIN REPEATS, METAL BINDING PROTEIN
1812	1ihb	A	129	304	1.6e-31	0.40	0.95		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A, B;	CELL CYCLE INHIBITOR P18-INK4C(INK6); CELL CYCLE INHIBITOR, P18-INK4C(INK6), ANKYRIN REPEAT, 2 CDK 4/6 INHIBITOR
1812	1ihb	A	167	314	4.8e-25	0.16	0.15		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A, B;	CELL CYCLE INHIBITOR P18-INK4C(INK6); CELL CYCLE INHIBITOR, P18-INK4C(INK6), ANKYRIN REPEAT, 2 CDK 4/6 INHIBITOR
1812	1ihb	A	75	229	1.1e-25	-0.11	0.63		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A, B;	CELL CYCLE INHIBITOR P18-INK4C(INK6); CELL CYCLE INHIBITOR, P18-INK4C(INK6), ANKYRIN REPEAT, 2 CDK 4/6 INHIBITOR
1812	1ikn	D	76	190	1.6e-27	0.07	-0.18		NF-KAPPA-B P65 SUBUNIT; CHAIN: A; NF-KAPPA-B P50D	TRANSCRIPTION FACTOR P65; P50D; TRANSCRIPTION

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									SUBUNIT; CHAIN: C; I-KAPPA-B-ALPHA; CHAIN: D;	FACTOR, IKB/NFKB COMPLEX
1812	1lkm	D	86	237	3.2e-32	0.31	0.71		NF-KAPPA-B P65 SUBUNIT; CHAIN: A; NF-KAPPA-B P50D SUBUNIT; CHAIN: C; I-KAPPA-B-ALPHA; CHAIN: D;	TRANSCRIPTION FACTOR P65; P50D; TRANSCRIPTION FACTOR, IKB/NFKB COMPLEX
1812	1lkm	D	94	300	1.3e-38	0.15	0.22		NF-KAPPA-B P65 SUBUNIT; CHAIN: A; NF-KAPPA-B P50D SUBUNIT; CHAIN: C; I-KAPPA-B-ALPHA; CHAIN: D;	TRANSCRIPTION FACTOR P65; P50D; TRANSCRIPTION FACTOR, IKB/NFKB COMPLEX
1812	1myo		162	271	1.4e-22	0.64	0.88		MYOTROPHIN; CHAIN: NULL	ANK-REPEAT MYOTROPHIN, ACETYLATION, NMR, ANK-REPEAT
1812	1nfi	E	76	237	1.6e-32	0.40	0.90		NF-KAPPA-B P65; CHAIN: A, C; NF-KAPPA-B P50; CHAIN: B, D; I-KAPPA-B-ALPHA; CHAIN: E, F;	COMPLEX (TRANSCRIPTION REG/ANK REPEAT) COMPLEX (TRANSCRIPTION REGULATION/ANK REPEAT), ANKYRIN 2 REPEAT HELIX
1812	1nfi	E	93	300	3.2e-38	0.18	0.43		NF-KAPPA-B P65; CHAIN: A, C; NF-KAPPA-B P50; CHAIN: B, D; I-KAPPA-B-ALPHA; CHAIN: E, F;	COMPLEX (TRANSCRIPTION REG/ANK REPEAT) COMPLEX (TRANSCRIPTION REGULATION/ANK REPEAT), ANKYRIN 2 REPEAT HELIX
1812	1yca	B	136	270	1.4e-22	0.41	0.99		P53; CHAIN: A; 53BP2; CHAIN: B;	COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS) P53BP2; ANKYRIN REPEATS, SH3, P53, TUMOR SUPPRESSOR, MULTIGENE 2

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										FAMILY, NUCLEAR PROTEIN, PHOSPHORYLATION, DISEASE MUTATION, 3 POLYMORPHISM, COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS)
1812	lycs	B	162	290	6.8e-22	0.47	0.90		P53; CHAIN: A; 53BP2; CHAIN: B;	COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS) P53BP2; ANKYRIN REPEATS, SH3, P53, TUMOR SUPPRESSOR, MULTIGENE 2 FAMILY, NUCLEAR PROTEIN, PHOSPHORYLATION, DISEASE MUTATION, 3 POLYMORPHISM, COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS)
1824	1zbd	B	57	107	0.0038	0.10	0.15		RAB-3A; CHAIN: A; RABPHILIN-3A; CHAIN: B;	COMPLEX (GTP-BINDING/EFFECTOR) RAS-RELATED PROTEIN RAB3A; COMPLEX (GTP-BINDING/EFFECTOR), G PROTEIN, EFFECTOR, RABCDR, 2 SYNAPTIC EXOCYTOSIS, RAB PROTEIN, RAB3A, RABPHILIN
1825	1f5w	A	19	123	0.0034	0.56	0.28		COXSACKIE VIRUS AND	VIRUS/VIRAL PROTEIN



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1825	1neu		19	119	3.4e-05	0.74	0.57		ADENOVIRUS RECEPTOR; CHAIN: A, B; MYELIN P0 PROTEIN; CHAIN: NULL;	RECEPTOR IMMUNOGLOBULIN V DOMAIN FOLD, SYMMETRIC DIMER STRUCTURAL PROTEIN MYELIN, STRUCTURAL PROTEIN, GLYCOPROTEIN, TRANSMEMBRANE, PHOSPHORYLATION, IMMUNOGLOBULIN FOLD, SIGNAL, MYELIN 2 MEMBRANE ADHESION MOLECULE
1827	1ac6	A	26	116	4.8e-36	0.01	0.76		T-CELL RECEPTOR ALPHA; CHAIN: A, B;	RECEPTOR RECEPTOR, V ALPHA DOMAIN, SITE-DIRECTED MUTAGENESIS, 2 THREE-DIMENSIONAL STRUCTURE, GLYCOPROTEIN, SIGNAL
1827	1fyt	D	30	124	9.6e-36	-0.19	0.96		HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR CHAIN: A; HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR-1 CHAIN: B; HEMAGGLUTININ HA1 PEPTIDE CHAIN: CHAIN: C; T-CELL RECEPTOR ALPHA CHAIN; CHAIN: D; T-CELL	IMMUNE SYSTEM HLA-DRI, DRA; HLA-DRI, DRB1 0101; TCR HA1.7 ALPHA CHAIN; TCR HA1.7 BETA CHAIN; PROTEIN-PROTEIN COMPLEX, IMMUNOGLOBULIN FOLD

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1827	1lcr	A	27	130	4.8e-37	-0.23	0.93		RECEPTOR BETA CHAIN; CHAIN: E;	RECEPTOR TCR; T-CELL, RECEPTOR, TRANSMEMBRANE, GLYCOPROTEIN, SIGNAL
1830	1awe		335	460	1.6e-17	0.45	-0.08		SOS1; CHAIN: NULL;	SIGNAL TRANSDUCTION SIGNAL TRANSDUCTION, SOS, PLECKSTRIN HOMOLOG (PH) DOMAIN
1830	1byl	A	98	268	3.4e-26	0.02	0.87		PIX; CHAIN: A;	TRANSPORT PROTEIN RHO-GTPASE EXCHANGE FACTOR, TRANSPORT PROTEIN
1830	1dbh	A	104	408	6.8e-26	0.22	0.94		HUMAN SOS 1; CHAIN: A;	GENE REGULATION SON OF SEVENLESS PROTEIN; GUANINE NUCLEOTIDE EXCHANGE FACTOR, GENE REGULATION
1830	1dbh	A	293	460	1.4e-18	0.07	0.12		HUMAN SOS 1; CHAIN: A;	GENE REGULATION SON OF SEVENLESS PROTEIN; GUANINE NUCLEOTIDE EXCHANGE FACTOR, GENE REGULATION
1830	1f5x	A	97	262	1e-22	0.58	1.00		RHO-GEF VAV; CHAIN: A;	SIGNALING PROTEIN 11 ALPHA-HELICES
1830	1fey	A	358	458	9.6e-14	0.04	-0.05		GRP1; CHAIN: A;	SIGNALING PROTEIN ARF1 GUANINE NUCLEOTIDE EXCHANGE FACTOR AND PH DOMAIN
1830	1rms		327	460	4.8e-18	0.23	0.11		SOS 1; CHAIN: NULL;	SIGNAL TRANSDUCTION SON

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1833	1c0t	A	73	313	3.2e-64	0.04	0.45		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1833	1c1c	B	74	313	3.2e-73	-0.10	0.36		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1833	1c9t	A	72	313	3.2e-68	-0.09	0.93		HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5' CHAIN: T; DNA (5' CHAIN: P;	TRANSFERASE/IMMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184LE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/IMMUNE 3 SYSTEM/DNA
1833	1c9t	B	72	313	4.8e-78	-0.11	0.46		HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT	TRANSFERASE/IMMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184LE, 3TC, PROTEIN-DNA 2 COMPLEX,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1833	1har		72	258	3.2e-56			61.15	REVERSE TRANSCRIPTASE HIV-1 REVERSE TRANSCRIPTASE (AMINO- TERMINAL HALF) (FINGERS IHAR 3 AND PALM SUBDOMAINS) (RT216) (E.C.2.7.7.49) IHAR 4	DRUG RESISTANCE, M184L, TRANSFERASE/IMMUNE 3 SYSTEM/DNA
1833	1har		72	258	3.2e-56	0.27	0.90		REVERSE TRANSCRIPTASE HIV-1 REVERSE TRANSCRIPTASE (AMINO- TERMINAL HALF) (FINGERS IHAR 3 AND PALM SUBDOMAINS) (RT216) (E.C.2.7.7.49) IHAR 4	
1833	1mm1		15	278	6.4e-49			121.40	MM1V REVERSE TRANSCRIPTASE; 1MML 4 CHAIN: NULL; 1MML 5	REVERSE TRANSCRIPTASE
1833	1mm1		77	277	6.4e-49	0.41	1.00		MM1V REVERSE TRANSCRIPTASE; 1MML 4 CHAIN: NULL; 1MML 5	REVERSE TRANSCRIPTASE
1833	1rth	A	72	313	8e-85	-0.02	0.80		HIV-1 REVERSE TRANSCRIPTASE; 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1833	1rth	B	72	313	4.8e-76	-0.19	0.21		HIV-1 REVERSE TRANSCRIPTASE; 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1833	1vrt	A	73	313	3.2e-84	-0.09	0.69		HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
1833	1vrt	B	74	313	4.8e-74	-0.10	0.15		HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
1833	3hvt	B	72	313	1.3e-68	-0.29	0.13		NUCLEOTIDYLTRANSFERASE REVERSE TRANSCRIPTASE (E.C.2.7.49) 3HVT 3	
1842	1d0s	A	24	266	6.8e-11	0.00	-0.20		NICOTINATE MONONUCLEOTIDE; 5,6- CHAIN: A;	TRANSFERASE DINUCLEOTIDE-BINDING MOTIF, PHOSPHORIBOSYL TRANSFERASE
1846	1ec4	A	12	136	1.7e-05	-0.21	0.12		KARYOPHERIN ALPHA; CHAIN: A, B; MYC PROTO-ONCOGENE PROTEIN; CHAIN: C, D, E, F;	TRANSPORT PROTEIN SERINE-RICH RNA POLYMERASE I SUPPRESSOR PROTEIN; ARM REPEAT
1846	1ec4	A	55	233	1.7e-11	-0.18	0.42		KARYOPHERIN ALPHA; CHAIN: A, B; MYC PROTO-ONCOGENE PROTEIN; CHAIN:	TRANSPORT PROTEIN SERINE-RICH RNA POLYMERASE I SUPPRESSOR PROTEIN; ARM

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1846	1ia1	A	5	181	3.1e-11	0.14	0.28		C, D, E, F; IMPORTIN ALPHA; CHAIN: A;	REPEAT NUCLEAR IMPORT RECEPTOR KARYOPHERIN ALPHA; NUCLEAR IMPORT RECEPTOR, NUCLEAR LOCALIZATION SIGNAL, 2 ARMADILLO REPEATS, AUTOINHIBITION, INTRASTERIC REGULATION
1846	3bet		61	282	6.8e-11	-0.10	0.43		BETA-CATENIN; CHAIN: NULL;	ARMADILLO REPEAT ARMADILLO REPEAT, BETA-CATENIN, CYTOSKELETON
1849	1dun		324	401	8e-11	-0.53	0.40		DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEODITHYDROLASE; CHAIN: NULL;	HYDROLASE DUTPASE, DUTP PYROPHOSPHATASE; HYDROLASE, DUTPASE, EIAY, TRIMERIC ENZYME, ASPARTYL PROTEASE
1849	1dun		332	401	1.7e-15	-0.91	0.45		DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEODITHYDROLASE; CHAIN: NULL;	HYDROLASE DUTPASE, DUTP PYROPHOSPHATASE; HYDROLASE, DUTPASE, EIAY, TRIMERIC ENZYME, ASPARTYL PROTEASE
1849	1euw	A	330	401	6.4e-08	-0.83	0.42		DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEODITHYDROLASE; CHAIN: A;	HYDROLASE DUTPASE, JELLY ROLL, MERCURY DERIVATIVE
1849	1f7d	A	335	401	1.4e-12	-0.88	0.34		POL. POLYPROTEIN; CHAIN: A, B;	VIRUS/VIRAL PROTEIN EIGHT STRANDED BETA-BARREL

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1849	1f7d	A	339	401	3.4e-14	-0.63	0.51		POL POLYPROTEIN; CHAIN: A, B;	VIRUS/VIRAL PROTEIN EIGHT STRANDED BETA-BARREL
1849	1f7r	A	335	401	1.4e-12	-0.75	0.47		POL POLYPROTEIN; CHAIN: A;	VIRUS/VIRAL PROTEIN EIGHT STRANDED BETA BARREL PROTEIN
1849	1f7r	A	339	401	3.4e-15	-0.78	0.76		POL POLYPROTEIN; CHAIN: A;	VIRUS/VIRAL PROTEIN EIGHT STRANDED BETA BARREL PROTEIN
1864	1bor		210	256	0.0035	-0.37	0.01		TRANSCRIPTION FACTOR PML; CHAIN: NULL;	TRANSCRIPTION REGULATION PROTO-ONCOGENE, NUCLEAR BODIES (PODS), LEUKEMIA, 2
1864	1bor		210	257	3.4e-10	-0.58	0.11		TRANSCRIPTION FACTOR PML; CHAIN: NULL;	TRANSCRIPTION REGULATION PROTO-ONCOGENE, NUCLEAR BODIES (PODS), LEUKEMIA, 2
1864	1chc		218	262	1.6e-07	-0.22	0.47		VIRUS EQUINE HERPES VIRUS-1 (C3HC4, OR RING DOMAIN) 1CHC 3 (NMR, 1 STRUCTURE) 1CHC 4	TRANSCRIPTION REGULATION
1864	1fbv	A	196	270	3.2e-09	-0.15	0.16		SIGNAL TRANSDUCTION PROTEIN CBL; CHAIN: A; ZAP-70 PEPTIDE; CHAIN: B; UBIQUITIN-CONJUGATING ENZYME B12-18 KDA UBCH7; CHAIN: C;	LIGASE CBL, UBCH7, ZAP-70, E2, UBIQUITIN, E3, PHOSPHORYLATION, 2 TYROSINE KINASE, UBIQUITINATION, PROTEIN DEGRADATION,
1864	1g25	A	217	257	1e-07	-0.57	0.43		CDK-ACTIVATING KINASE	METAL BINDING PROTEIN RING

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEOFOLD score	Compound	PDB annotation
1864	1rmd		201	304	4.8e-05	-0.67	0.05		ASSEMBLY FACTOR MAT1; CHAIN: A;	FINGER PROTEIN MAT1; RING FINGER (C3HC4)
									RAG1; CHAIN: NULL;	DNA-BINDING PROTEIN V(D)J RECOMBINATION ACTIVATING PROTEIN 1; RAG1, V(D)J RECOMBINATION, ANTIBODY, MAD, RING FINGER, 2 ZINC NUCLEAR CLUSTER, ZINC FINGER, DNA-BINDING PROTEIN
1864	1rmd		217	257	1.4e-09	-0.58	0.82		RAG1; CHAIN: NULL;	DNA-BINDING PROTEIN V(D)J RECOMBINATION ACTIVATING PROTEIN 1; RAG1, V(D)J RECOMBINATION, ANTIBODY, MAD, RING FINGER, 2 ZINC NUCLEAR CLUSTER, ZINC FINGER, DNA-BINDING PROTEIN
1887	1bd2	E	22	138	6.4e-47	0.01	0.34		HLA-A 0201; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; TAX PEPTIDE; CHAIN: C; T CELL RECEPTOR ALPHA; CHAIN: D; T CELL RECEPTOR BETA; CHAIN: E;	COMPLEX (MHC/VTRAL PEPTIDE/RECEPTOR) HLA A2 HEAVY CHAIN; COMPLEX (MHC/VTRAL PEPTIDE/RECEPTOR)
1887	1bec		23	138	9.6e-48	0.05	0.12		14.3.D T CELL ANTIGEN RECEPTOR; 1BEC 5 CHAIN: NULL; 1BEC 6	RECEPTOR T CELL RECEPTOR 1BEC 14



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1887	1bwmm	A	23	177	1.6e-55	-0.12	0.22		ALPHA-BETA T CELL RECEPTOR (TCR) (D10); CHAIN: A;	IMMUNE SYSTEM IMMUNOGLOBULIN, IMMUNORECEPTOR, IMMUNE SYSTEM
1887	1nfd	B	20	138	1.6e-45	0.48	1.00		N15 ALPHA-BETA T-CELL RECEPTOR; CHAIN: A, B, C, D; H57 FAB; CHAIN: E, F, G, H	COMPLEX (IMMUNORECEPTOR/IMMUNOGLOBULIN) COMPLEX (IMMUNORECEPTOR/IMMUNOGLOBULIN)
1887	1nfd	B	20	183	1.6e-45			74.63	N15 ALPHA-BETA T-CELL RECEPTOR; CHAIN: A, B, C, D; H57 FAB; CHAIN: E, F, G, H	COMPLEX (IMMUNORECEPTOR/IMMUNOGLOBULIN) COMPLEX (IMMUNORECEPTOR/IMMUNOGLOBULIN)
1887	2fb4	H	31	179	4.8e-15	0.03	-0.05		IMMUNOGLOBULIN IMMUNOGLOBULIN FAB 2FB4 4	
1889	1a9n	C	44	101	3.4e-06	-0.14	0.21		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A'; CHAIN: A, C; U2 B"; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA, RNA, SNRNP, RIBONUCLEOPROTEIN
1889	1dce	A	43	102	6.8e-05	-0.49	0.19		RAB GERANYLGERANYLTRANSFERASE ALPHA SUBUNIT; CHAIN: A, C; RAB GERANYLGERANYLTRANSFERASE BETA SUBUNIT; CHAIN:	TRANSFERASE CRYSTAL STRUCTURE, RAB GERANYLGERANYLTRANSFERASE, 2.0 Å 2 RESOLUTION, N-FORMYL METHIONINE, ALPHA SUBUNIT, BETA SUBUNIT

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1895	1b9f	A	171	205	1e-05	-0.77	0.13		INTEGRASE; CHAIN: A;	TRASFERRASE DNA INTEGRATION, TRASFERRASE
1895	1c0t	A	5	154	1.6e-30	0.21	-0.05		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRASFERRASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1895	1c9r	A	5	175	3.2e-31	0.13	-0.13		HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5'-CHAIN: T; DNA (5'-CHAIN: P;	TRASFERRASE/MMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184IE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRASFERRASE/MMUNE 3 SYSTEM/DNA
1895	1l21	A	34	183	4.8e-33	-0.05	0.04		RIBONUCLEASE H; CHAIN: A;	HYDROLASE RNASE H, NUCLEASE, RNASE H*, RIBONUCLEASE H, METAL-BINDING 2 PROTEIN, PROTEIN FOLDING
1895	1rl1		36	161	9.6e-27	-0.18	0.16		HYDROLASE/ENDORIBONUCLEASE) RIBONUCLEASE H (E.C.3.1.26.4) 1RL1.3	
1895	1rl1		76	170	1e-17	0.04	0.68		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1895	1rth	A	5	156	4.8e-31	0.02	-0.11		(E.C.3.1.26.4) IRL 3 HIV-1 REVERSE TRANSCRIPTASE, 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1895	1vrt	A	5	154	4.8e-26	0.01	-0.08		HIV-1 REVERSE TRANSCRIPTASE, 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
1911	1cod		46	116	3.2e-17	-0.58	0.19		SHORT NEUROTOXIN COBROTOXIN (NMR, AVERAGE STRUCTURE) 1CODA 2	
1911	1kba	A	41	117	1.3e-18	-0.28	0.01		TOXIN KAPPA- BUNGAROTOXIN 1KBA 3	
1911	1nea		41	116	6.4e-19	-0.22	0.03		TOXIN TOXIN ALPHA (NMR, 8 STRUCTURES) 1NEA 3	
1911	2abx	A	46	117	1.6e-21	0.02	0.30		POSTSYNAPTIC NEUROTOXIN ALPHA-*BUNGAROTOXIN 2ABX 4	
1912	1bdq	A	37	129	0.00031	0.47	0.70		HIV-1 PROTEASE; CHAIN: A, B;	HYDROLASE HYDROLASE, AIDS, POLYPROTEIN, ASPARTYL PROTEASE, ACID 2 PROTEASE, HYDROXYETHYLENE ISOSTERE INHIBITOR,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1912	1bdd	A	43	129	0.0085	0.43	0.90		HIV-1 PROTEASE; CHAIN: A, B;	SUBSTRATE 3 ANALOGUE INHIBITOR
1912	1bwb	A	37	129	0.00068	0.13	0.45		HIV-1 PROTEASE; CHAIN: A, B;	HYDROLASE HIV-1 PROTEASE, HYDROLASE
1912	1daz	C	37	132	0.00024	0.59	0.89		PEPTIDE INHIBITOR; CHAIN: A, B; HIV-1 PROTEASE (RETROPEPSIN); CHAIN: C, D;	HYDROLASE HIV-1 PROTEASE, HYDROLASE, MUTANT, DIMER, INHIBITOR, OCCUPANCY
1912	1hvc		37	129	0.0068	0.10	0.29		HYDROLASE(ACID PROTEASE) HIV-1 PROTEASE (TETHERED DIMER LINKED BY 1HVC3 GLY-SER-SER-GLY) COMPLEXED WITH A-76928 1HVC 4	
1912	1ida	A	37	129	0.00068	0.27	0.35		HYDROLASE(ACID PROTEINASE) HUMAN IMMUNODEFICIENCY VIRUS TYPE 2 (HIV-2) PROTEASE 1IDA 3 COMPLEXED WITH THE INHIBITOR 1ILA 1906 CONTAINING THE 1IDA 4	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1912	1mt	A	37	132	0.00017	0.59	0.45		HYDROXYETHYLAMINE DIPEPTIDE ISOSTERE 1DA 5 HIV-1 PROTEASE; A CYCLIC PHE-ILE-VAL PEPTIDOMIMETIC INHIBITOR; CHAIN: C;	COMPLEX (ASPARTYL PROTEASE/INHIBITOR) HIV-1 PR: HYDROLASE, ASPARTYL PROTEINASE, AIDS, PEPTIDE, INHIBITOR
1912	1sip		37	132	0.0068	0.80	0.54		HYDROLASE/ACID PROTEINASE) SIMIAN IMMUNODEFICIENCY VIRUS (SIV) PROTEINASE 1SP 3 (SIV MAC251-32H ISOLATE) (E.C.3.4.23.-) 1SP 4	
1913	1b0x	A	111	170	1e-05	0.43	0.64		EPHA4 RECEPTOR TYROSINE KINASE; CHAIN: A;	TRANSFERASE RECEPTOR TYROSINE KINASE, PROTEIN INTERACTION MODULE, 2 DIMERIZATION DOMAIN, TRANSFERASE
1913	1b0x	A	277	347	3.2e-15	0.33	0.10		EPHA4 RECEPTOR TYROSINE KINASE; CHAIN: A;	TRANSFERASE RECEPTOR TYROSINE KINASE, PROTEIN INTERACTION MODULE, 2 DIMERIZATION DOMAIN, TRANSFERASE
1913	1b4f	A	107	171	2.7e-11	0.53	0.81		EPHB2; CHAIN: A, B, C, D, E, F, G, H;	SIGNAL TRANSDUCTION SAM DOMAIN, EPH RECEPTOR, SIGNAL TRANSDUCTION, OLIGOMER

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1913	1b4f	A	182	249	9.6e-15	0.14	0.37		EPHB2; CHAIN: A, B, C, D, E, F, G, H;	SIGNAL TRANSDUCTION SAM DOMAIN, EPH RECEPTOR, SIGNAL TRANSDUCTION, OLIGOMER
1913	1b4f	A	270	348	6.4e-19	0.28	0.13		EPHB2; CHAIN: A, B, C, D, E, F, G, H;	SIGNAL TRANSDUCTION SAM DOMAIN, EPH RECEPTOR, SIGNAL TRANSDUCTION, OLIGOMER
1913	1sge		111	171	3.4e-06	0.58	0.77		EPHRII TYPE-B RECEPTOR 2; CHAIN: NULL;	TYROSINE-PROTEIN KINASE NMR, RECEPTOR OLIGOMERIZATION, EPH RECEPTORS, TYROSINE 2 PHOSPHORYLATION, SIGNAL TRANSDUCTION, TYROSINE-PROTEIN 3 KINASE
1913	1sge		184	249	8e-14	0.34	0.52		EPHRII TYPE-B RECEPTOR 2; CHAIN: NULL;	TYROSINE-PROTEIN KINASE NMR, RECEPTOR OLIGOMERIZATION, EPH RECEPTORS, TYROSINE 2 PHOSPHORYLATION, SIGNAL TRANSDUCTION, TYROSINE-PROTEIN 3 KINASE
1914	1be9	A	19	69	3.2e-15	-0.21	0.29		PSD-95; CHAIN: A; CRIFT; CHAIN: B;	PEPTIDE RECOGNITION PEPTIDE RECOGNITION, PROTEIN LOCALIZATION
1914	1pdr		19	62	8e-12	0.08	0.04		HUMAN DISCS LARGE PROTEIN; CHAIN: NULL;	SIGNAL TRANSDUCTION HDLG, DHR3 DOMAIN; SIGNAL

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1914	1qav	A	19	50	1.3e-09	-0.08	0.66		ALPHA-1 SYNTROPHIN (RESIDUES 77-171); CHAIN: A; NEURONAL NITRIC OXIDE SYNTHASE (RESIDUES 1-130); CHAIN: B;	TRANSDUCTION, SH3 DOMAIN, REPEAT MEMBRANE PROTEIN/OXIDOREDUCTASE BETA-FINGER, HETERODIMER
1914	3pdc	A	19	48	9.6e-06	-0.15	0.46		TYROSINE PHOSPHATASE (PTP-BAS, TYPE 1); CHAIN: A;	HYDROLASE PDZ DOMAIN, HUMAN PHOSPHATASE, HPTP1E, PTP-BAS, SPECIFICITY 2 OF BINDING
1920	1b8q	A	136	171	1.7e-06	-0.86	0.25		NEURONAL NITRIC OXIDE SYNTHASE; CHAIN: A; HEPTAPEPTIDE; CHAIN: B; PSD-95; CHAIN: A; CRPT; CHAIN: B;	OXIDOREDUCTASE PDZ DOMAIN, NNOS, NITRIC OXIDE SYNTHASE
1920	1be9	A	136	188	1.4e-06	-0.42	0.41		HEPTAPEPTIDE; CHAIN: B;	PEPTIDE RECOGNITION PEPTIDE RECOGNITION, PROTEIN LOCALIZATION
1920	1kwa	A	136	170	1.4e-05	-0.86	0.45		HCASK/LIN-2 PROTEIN; CHAIN: A, B;	KINASE HCASK, G1/GF REPEAT, DHR, PDZ DOMAIN, NEUREXIN, SYNDECAN, RECEPTOR CLUSTERING, KINASE
1920	1qau	A	136	171	0.0001	-0.73	0.99		NEURONAL NITRIC OXIDE SYNTHASE (RESIDUES 1-130); CHAIN: A;	OXIDOREDUCTASE BETA-FINGER
1920	1qav	A	136	170	3.4e-05	-0.55	0.71		ALPHA-1 SYNTROPHIN (RESIDUES 77-171); CHAIN: A; NEURONAL NITRIC OXIDE	MEMBRANE PROTEIN/OXIDOREDUCTASE BETA-FINGER, HETERODIMER

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1920	1qlc	A	135	170	1.4e-05	-0.72	0.94		SYNTHASE (RESIDUES 1-130); CHAIN: B;	
1920	3pdz	A	129	170	0.0037	-0.79	0.93		POSTSYNAPTIC DENSITY PROTEIN 95; CHAIN: A;	PEPTIDE RECOGNITION PSD-95; PDZ DOMAIN, NEURONAL NITRIC OXIDE SYNTHASE, NMDA RECEPTOR 2 BINDING
									TYROSINE PHOSPHATASE (PTP-BAS, TYPE 1); CHAIN: A;	HYDROLASE PDZ DOMAIN, HUMAN PHOSPHATASE, HPTPIE, PTP-BAS, SPECIFICITY 2 OF BINDING
1930	1ad0	A	29	146	3.2e-65	0.38	1.00		FAB FRAGMENT, ANTIBODY A5B7; CHAIN: A, B, C, D;	IMMUNOGLOBULIN IMMUNOGLOBULIN, FAB FRAGMENT
1930	1b2w	L	28	146	9.6e-69	0.30	0.98		ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H;	IMMUNE SYSTEM IMMUNOGLOBULIN; IMMUNOGLOBULIN ANTIBODY ENGINEERING, HUMANIZED AND CHIMERIC ANTIBODY, FAB, 2 X-RAY STRUCTURE, THREE-DIMENSIONAL STRUCTURE, GAMMA-3 INTERFERON, IMMUNE SYSTEM
1930	1b6d	A	28	146	1.1e-69	0.45	0.98		IMMUNOGLOBULIN; CHAIN: A, B;	IMMUNOGLOBULIN IMMUNOGLOBULIN, KAPPA LIGHT-CHAIN DIMER HEADER
1930	1bj1	L	28	146	1.4e-71	0.40	0.99		FAB FRAGMENT; CHAIN: L, H, J, K; VASCULAR	COMPLEX (ANTIBODY/ANTIGEN) FAB-12;



Table 5

SEQ ID NO.	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1930	1bvk	A	27	133	1.6e-59			62.65	ENDOTHELIAL GROWTH FACTOR; CHAIN: V, W;	VEGF: COMPLEX (ANTIBODY/ANTIGEN), ANGIOGENIC FACTOR
1930	1ce1	L	28	146	1.3e-68	0.40	0.99		HULYS11; CHAIN: A, B, D, E, LYSOZYME; CHAIN: C, F;	COMPLEX (HUMANIZED ANTIBODY/HYDROLASE) MURAMIDASE; HUMANIZED ANTIBODY, ANTIBODY COMPLEX, FV, ANTI-LYSOZYME, 2 COMPLEX (HUMANIZED ANTIBODY/HYDROLASE)
1930	1dec	A	28	146	1.6e-72	0.63	0.99		CAMPATH-1H; LIGHT CHAIN; CHAIN: L; CAMPATH-1H; HEAVY CHAIN; CHAIN: H; PEPTIDE ANTIGEN; CHAIN: P;	ANTIBODY THERAPEUTIC, ANTIBODY, CD52
1930	1dfb	L	28	146	9.6e-69	0.47	0.98		IGM RF 2A2; CHAIN: A, C, E; IGM RF 2A2; CHAIN: B, D, F; IMMUNOGLOBULIN G BINDING PROTEIN A; CHAIN: G, H;	IMMUNE SYSTEM FAB-IBP COMPLEX CRYSTAL STRUCTURE 2.7A RESOLUTION BINDING 2 OUTSIDE THE ANTIGEN COMBINING SITE SUPERANTIGEN FAB VH3 3 SPECIFICITY
1930	1fvd	A	28	146	3.2e-70	0.49	0.68		IMMUNOGLOBULIN 3D6 FAB IDFB 3	
									IMMUNOGLOBULIN FAB FRAGMENT OF HUMANIZED ANTIBODY 4D5, VERSION 4 IFVD 3	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1930	1vge	L	29	146	1.6e-68	0.51	1.00		TRI.9 FAB; CHAIN: L, H <sub>1</sub>	IMMUNOGLOBULIN TRI.9, ANTI-THYROID PEROXIDASE, AUTOANTIBODY, 2
1930	2fgw	L	28	146	8e-72	0.57	1.00		IMMUNOGLOBULIN FAB FRAGMENT OF A HUMANIZED VERSION OF THE ANTI-CD18 2FGW 3 ANTIBODY T52' (HUH52-OZ FAB) 2FGW 4	IMMUNOGLOBULIN
1934	1d1d	A	9	69	0.0062	-0.64	0.22		CAPSID PROTEIN; CHAIN: A <sub>1</sub>	VIRUS/VIRAL PROTEIN TWO INDEPENDENT DOMAINS HELICAL BUNDLES, VIRUS/VIRAL PROTEIN
1935	1b9f	A	8	150	1.4e-25	0.23	0.82		INTEGRASE; CHAIN: A <sub>1</sub>	TRASFERASE DNA INTEGRATION, TRASFERASE
1935	1b13	C	8	150	4.8e-28	0.25	0.55		INTEGRASE; CHAIN: A, B, C <sub>1</sub>	DNA INTEGRATION DNA INTEGRATION, AIDS, POLYPROTEIN, HYDROLASE, 2 ENDONUCLEASE, POLYNUCLEOTIDYL TRANSFERASE, DNA BINDING 3 (VIRAL)
1935	1c0m	A	1	132	3.4e-23	0.17	0.86		INTEGRASE; CHAIN: A, B, C <sub>1</sub> , D <sub>1</sub>	TRASFERASE INTEGRASE, ROUS SARCOMA VIRUS, HIV, X-RAY CRYSTALLOGRAPHY, 2

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1935	1c0m	A	2	126	3.2e-20	0.04	0.58		INTEGRASE; CHAIN: A, B, C, D;	PROTEIN STRUCTURE, TRANSFERASE
1935	1c1a	B	1	131	1.4e-20	0.28	0.92		RSV INTEGRASE; CHAIN: A, B;	VIRUS/VIRAL PROTEIN INTEGRASE, ROUS SARCOMA VIRUS, HIV, X-RAY CRYSTALLOGRAPHY, 2
1935	1c1a	B	2	131	1.6e-20	0.19	0.96		RSV INTEGRASE; CHAIN: A, B;	VIRUS/VIRAL PROTEIN INTEGRASE, ROUS SARCOMA VIRUS, HIV, X-RAY CRYSTALLOGRAPHY, 2
1935	1c29	A	2	120	3.4e-21	0.17	0.86		AVIAN SARCOMA VIRUS INTEGRASE; CHAIN: A;	TRANSFERASE MIXED BETA-SHEET SURROUNDED BY ALPHA-HELICES
1935	1cx4	A	8	166	6.4e-24	-0.09	0.45		INTEGRASE; CHAIN: A, B;	VIRUS/VIRAL PROTEIN SH3-LIKE DOMAIN, NONSPECIFIC DNA BINDING BETA SHEET, CIS-2 PROLINE
1935	1cxq	A	8	150	6.4e-21	0.04	0.40		POLYPROTEIN; CHAIN: A, B;	VIRUS/VIRAL PROTEIN HIV-1 INTEGRASE, POLYNUCLEOTIDYL TRANSFERASE, DNA-BINDING 2

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1935	1q64	A	8	150	3.2e-23	-0.09	0.63		HIV-1 INTEGRASE; CHAIN: A, B, C;	PROTEIN, DD35E HYDROLASE DNA INTEGRATION, INTEGRASE, HIV, HYDROLASE, ASPARTYL 2 PROTEASE, ENDONUCLEASE
1940	1a0b		148	192	9.6e-09	-0.02	0.58		HIV-2 INTEGRASE; CHAIN: NULL;	INTEGRASE INTEGRASE, AIDS, POLYPROTEIN
1940	1c0t	A	1	125	1.4e-20	-0.30	0.00		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON- NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1940	1h0h	A	21	123	3.2e-15	0.00	0.41		HYDROLASE(ENDORIBONUC LEASE) RIBONUCLEASE H DOMAIN OF /HIV-1\$ REVERSE TRANSCRIPTASE 1HRH 3	
1940	1h0h	A	22	152	3.4e-17	0.07	0.17		HYDROLASE(ENDORIBONUC LEASE) RIBONUCLEASE H DOMAIN OF /HIV-1\$ REVERSE TRANSCRIPTASE 1HRH 3	
1940	1r1l		24	156	6.8e-14	0.10	0.39		HYDROLASE(ENDORIBONUC LEASE) RIBONUCLEASE H (E.C.3.1.26.4) 1R1L 3	
1940	1vrt	A	1	125	4.8e-18	-0.11	0.13		HIV-1 REVERSE TRANSCRIPTASE, 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1952	1d0b	A	157	203	1.3e-05	-0.01	0.93		INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION
1952	1d0b	A	74	200	1.3e-22	0.37	0.05		INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION
1952	1dce	A	71	149	3.2e-10	0.24	-0.13		RAB GERANYLGERANYLTRANSFERASE ALPHA SUBUNIT; CHAIN: A, C; RAB GERANYLGERANYLTRANSFERASE BETA SUBUNIT; CHAIN: B, D;	TRANSFERASE CRYSTAL STRUCTURE, RAB GERANYLGERANYLTRANSFERASE, 2.0 A 2 RESOLUTION, N-FORMYLMETHIONINE, ALPHA SUBUNIT, BETA SUBUNIT
1952	1dce	A	97	197	1.1e-10	-0.11	0.22		RAB GERANYLGERANYLTRANSFERASE ALPHA SUBUNIT; CHAIN: A, C; RAB GERANYLGERANYLTRANSFERASE BETA SUBUNIT; CHAIN: B, D;	TRANSFERASE CRYSTAL STRUCTURE, RAB GERANYLGERANYLTRANSFERASE, 2.0 A 2 RESOLUTION, N-FORMYLMETHIONINE, ALPHA SUBUNIT, BETA SUBUNIT
1952	1ds9	A	70	191	4.8e-09	0.10	0.21		OUTER ARM DYNEIN; CHAIN: A;	CONTRACTILE PROTEIN LEUCINE-RICH REPEAT, BETA-BETA-ALPHA CYLINDER, DYNEIN, 2 CHILAMYDOMONAS, FLAGELLA
1952	1ds9	A	99	197	6.4e-13	-0.47	0.25		OUTER ARM DYNEIN; CHAIN: A;	CONTRACTILE PROTEIN LEUCINE-RICH REPEAT, BETA-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										BETA-ALPHA CYLINDER, DYNEIN, 2 CHLAMYDOMONAS, FLAGELLA
1953	1ayz	A	78	218	9.6e-46	0.35	0.98		UBIQUITIN-CONJUGATING ENZYME RAD6; CHAIN: A, B, C;	UBIQUITIN CONJUGATION UBC2; UBIQUITIN CONJUGATION, UBIQUITIN-CONJUGATING ENZYME
1953	1ayz	A	78	220	9.6e-46			54.97	UBIQUITIN-CONJUGATING ENZYME RAD6; CHAIN: A, B, C;	UBIQUITIN CONJUGATION UBC2; UBIQUITIN CONJUGATION, UBIQUITIN-CONJUGATING ENZYME
1953	1c4z	D	82	218	1.3e-35	0.17	0.83		UBIQUITIN-PROTEIN LIGASE E3A; CHAIN: A, B, C; UBIQUITIN CONJUGATING ENZYME E2; CHAIN: D;	LIGASE E6AP; UBCH7; BILOBAL STRUCTURE, ELONGATED SHAPE, E3 UBIQUITIN LIGASE, E2 2 UBIQUITIN CONJUGATING ENZYME
1953	1c4z	D	82	220	1.3e-35			59.93	UBIQUITIN-PROTEIN LIGASE E3A; CHAIN: A, B, C; UBIQUITIN CONJUGATING ENZYME E2; CHAIN: D;	LIGASE E6AP; UBCH7; BILOBAL STRUCTURE, ELONGATED SHAPE, E3 UBIQUITIN LIGASE, E2 2 UBIQUITIN CONJUGATING ENZYME
1953	1qcq	A	77	217	4.8e-50	0.15	0.99		UBIQUITIN CONJUGATING ENZYME; CHAIN: A;	LIGASE UBIQUITIN, UBIQUITIN-CONJUGATING ENZYME, YEAST
1953	1qcq	A	79	220	4.8e-50			55.68	UBIQUITIN CONJUGATING ENZYME; CHAIN: A;	LIGASE UBIQUITIN, UBIQUITIN-CONJUGATING ENZYME, YEAST

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1953	1u9a	A	77	219	3.2e-45	0.01	0.37		UBC9; CHAIN: NULL;	UBIQUITIN-CONJUGATING ENZYME UBIQUITIN-CONJUGATING ENZYME; UBIQUITIN-CONJUGATING ENZYME, UBIQUITIN-DIRECTED 2 PROTEOLYSIS, CELL CYCLE CONTROL, LIGASE
1953	2aak		77	218	1.6e-49	0.17	0.58		UBIQUITIN CONJUGATING ENZYME; CHAIN: NULL;	UBIQUITIN CONJUGATION UBC1; UBIQUITIN CONJUGATION, LIGASE
1953	2aak		77	220	1.6e-49			55.30	UBIQUITIN CONJUGATING ENZYME; CHAIN: NULL;	UBIQUITIN CONJUGATION UBC1; UBIQUITIN CONJUGATION, LIGASE
1953	2e2c		69	220	6.4e-43			57.76	UBIQUITIN CONJUGATING ENZYME; CHAIN: NULL;	UBIQUITIN CONJUGATION UBIQUITIN CONJUGATION, UBIQUITIN CARRIER PROTEIN, THIOESTER 2 BOND, LIGASE
1953	2e2c		76	219	6.4e-43	0.24	0.81		UBIQUITIN CONJUGATING ENZYME; CHAIN: NULL;	UBIQUITIN CONJUGATION UBIQUITIN CONJUGATION, UBIQUITIN CARRIER PROTEIN, THIOESTER 2 BOND, LIGASE
1953	2uc2		78	219	6.4e-43	0.29	0.66		UBIQUITIN CONJUGATING ENZYME; CHAIN: NULL;	UBIQUITIN CONJUGATION UBC7; UBIQUITIN CONJUGATION, LIGASE, YEAST
1954	1ek7	A	31	435	0	0.92	1.00		GELATINASE A; CHAIN: A;	HYDROLASE MMP-2, 72KD TYPE IV COLLAGENASE; HYDROLASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1954	1ck7	A	31	445	0			536.83	GELATINASE A; CHAIN: A;	(METALLOPROTEINASE), FULL-LENGTH, METALLOPROTEINASE, 2 GELATINASE A
										HYDROLASE MMP-2, 72KD TYPE IV COLLAGENASE; HYDROLASE (METALLOPROTEINASE), FULL-LENGTH, METALLOPROTEINASE, 2 GELATINASE A
1954	1cxw	A	275	334	3.4e-28			108.03	HUMAN MATRIX METALLOPROTEINASE 2; CHAIN: A;	HYDROLASE COL-2; BETA SHEET, ALPHA HELIX, HYDROLASE
1954	1cxw	A	276	334	1.6e-25	1.37	1.00		HUMAN MATRIX METALLOPROTEINASE 2; CHAIN: A;	HYDROLASE COL-2; BETA SHEET, ALPHA HELIX, HYDROLASE
1954	1cxw	A	276	334	3.4e-28	1.37	1.00		HUMAN MATRIX METALLOPROTEINASE 2; CHAIN: A;	HYDROLASE COL-2; BETA SHEET, ALPHA HELIX, HYDROLASE
1958	1bbd	A	207	281	3.2e-27	-0.01	0.07		RAB-3A; CHAIN: A; RABPHILIN-3A; CHAIN: B;	COMPLEX (GTP-BINDING/EFFECTOR) RAS-RELATED PROTEIN RAB3A; COMPLEX (GTP-BINDING/EFFECTOR), G PROTEIN, EFFECTOR, RABCDR, 2 SYNAPTIC EXOCYTOSIS, RAB



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1958	2ngr	A	211	283	3.2e-23	0.04	-0.12		GTP BINDING PROTEIN (G25K); CHAIN: A; GTPASE ACTIVATING PROTEIN (RHG); CHAIN: B;	PROTEIN, RAB3A, RABPHILIN HYDROLASE CDC42/CDC42GAP; CDC42/CDC42GAP; TRANSITION STATE, G-PROTEIN, GAP, CDC42, ALF3, HYDROLASE
1958	3rab	A	210	281	3.2e-28	0.01	-0.01		RAB3A; CHAIN: A;	HYDROLASE G PROTEIN, VESICULAR TRAFFICKING, GTP HYDROLYSIS, RAB 2 PROTEIN, NEUROTRANSMITTER RELEASE, HYDROLASE
1965	1c0t	A	12	144	3.2e-39	-0.66	0.19		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1965	1c1c	B	12	144	1.1e-41	-0.35	0.19		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1965	1c9t	A	12	144	4.8e-48	-0.48	0.36		HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H;	TRANSFERASE/IMMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184ILE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/IMMUNE 3

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1965	1c9r	B	12	144	1.1e-46	-0.37	0.07		DNA (5'-CHAIN: T; DNA (5'-CHAIN: P; HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5'-CHAIN: T; DNA (5'-CHAIN: P; REVERSE TRANSCRIPTASE HIV-1 REVERSE TRANSCRIPTASE (AMINO-TERMINAL HALF) (FINGERS 1HAR 3 AND PALM SUBDOMAINS) (RT216) (E.C.2.7.7.49) 1HAR 4	SYSTEM/DNA TRANSFERASE/IMMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184ILE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/IMMUNE 3 SYSTEM/DNA
1965	1har		12	103	4.8e-36	-0.43	0.25		REVERSE TRANSCRIPTASE HIV-1 REVERSE TRANSCRIPTASE (AMINO-TERMINAL HALF) (FINGERS 1HAR 3 AND PALM SUBDOMAINS) (RT216) (E.C.2.7.7.49) 1HAR 4	
1965	1rth	A	12	144	9.6e-46	-0.49	0.23		HIV-1 REVERSE TRANSCRIPTASE, 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1965	1rth	B	12	144	9.6e-42	-0.41	0.18		HIV-1 REVERSE TRANSCRIPTASE, 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1965	1vrt	A	12	144	9.6e-46	-0.56	0.13		HIV-1 REVERSE TRANSCRIPTASE, 1VRT 4	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
1965	1vrt	B	12	144	9.6e-42	-0.38	0.00		CHAIN: A, B; 1VRT 5	REVERSE TRANSCRIPTASE 1VRT 15
									HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1
1965	3hvt	B	12	126	1.6e-38	-0.20	0.13		CHAIN: A, B; 1VRT 5	REVERSE TRANSCRIPTASE 1VRT 15
									NUCLEOTIDYLTRANSFERASE REVERSE TRANSCRIPTASE (E.C.2.7.7.49) 3HVT 3	
1966	1c0t	A	25	205	1.6e-51	-0.12	0.80		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1966	1c1c	B	38	205	8e-47	-0.11	0.17		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
1966	1c9r	A	13	206	1.1e-50	-0.46	0.36		HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5'-CHAIN: T; DNA (5'-	TRANSFERASE/IMMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184LE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/IMMUNE 3 SYSTEM/DNA

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1966	1c9t	B	59	201	4.8e-52	-0.22	0.62		CHAIN: P; HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5'- CHAIN: T; DNA (5'- CHAIN: P;	TRANSFERASE/MIMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184ILE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/MIMUNE 3 SYSTEM/DNA
1966	1har		1	182	6.4e-38			50.80	REVERSE TRANSCRIPTASE HIV-1 REVERSE TRANSCRIPTASE (AMINO- TERMINAL HALF) (FINGERS IHAR 3 AND PALM SUBDOMAINS) (RT216) (E.C.2.7.7.49) IHAR 4	
1966	1har		72	177	6.4e-38	-0.20	0.82		REVERSE TRANSCRIPTASE HIV-1 REVERSE TRANSCRIPTASE (AMINO- TERMINAL HALF) (FINGERS IHAR 3 AND PALM SUBDOMAINS) (RT216) (E.C.2.7.7.49) IHAR 4	
1966	1mm1		1	197	9.6e-37			92.17	MM1V REVERSE TRANSCRIPTASE, 1MML 4 CHAIN: NULL; 1MML 5 (E.C.2.7.7.49) IHAR 4	REVERSE TRANSCRIPTASE
1966	1rth	A	38	205	3.2e-54	-0.33	0.89		HIV-1 REVERSE	NUCLEOTIDYLTRANSFERASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Pst Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1966	1rth	B	38	205	3.2e-49	-0.03	0.24		TRANSCRIPTASE; 1RTH 4 CHAIN: A, B; 1RTH 5	HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1966	1vrt	A	38	205	3.2e-54	0.08	0.93		HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
1966	1vrt	B	38	205	8e-48	-0.10	0.29		HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
1966	3hvt	B	38	183	1.6e-48	-0.26	0.49		NUCLEOTIDYLTRANSFERASE REVERSE TRANSCRIPTASE (E.C.2.7.7.49) 3HVT 3	
1973	1a0p		1	152	4.8e-38	0.22	0.98		SITE-SPECIFIC RECOMBINASE XERD; CHAIN: NULL;	DNA RECOMBINATION XERD, RECOMBINASE, DNA BINDING, DNA RECOMBINATION
1973	1ae9	A	1	129	1.4e-15	0.09	0.45		LAMBDA INTEGRASE; CHAIN: A, B;	DNA RECOMBINATION DNA RECOMBINATION, INTEGRASE, SITE-SPECIFIC RECOMBINATION
1973	1ae9	B	1	136	1.6e-16	0.24	0.27		LAMBDA INTEGRASE; CHAIN: A, B;	DNA RECOMBINATION DNA RECOMBINATION, INTEGRASE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1973	1a1h	A	2	139	8e-22	0.11	-0.05		HP1 INTEGRASE; CHAIN: A, B, C, D;	SITE-SPECIFIC RECOMBINATION DNA INTEGRATION, RECOMBINATION
1973	2cix	A	1	139	3.2e-14	0.20	-0.07		CRE RECOMBINASE; CHAIN: A, B; DNA; CHAIN: C, D;	COMPLEX (RECOMBINASE/DNA) CRE-H12; CRE RECOMBINASE, HOLLIDAY JUNCTION, RECOMBINATION, 2 COMPLEX (RECOMBINASE/DNA)
1973	4cix	A	1	139	9.6e-12	0.44	-0.01		CRE RECOMBINASE; CHAIN: A, B; DNA (35 NUCLEOTIDE CRE RECOGNITION SITE); CHAIN: C, D;	PROTEIN/DNA CRE RECOMBINASE, DNA BENDING, RECOMBINATION, PROTEIN-DNA 2 INTERACTION, PROTEIN/DNA
1973	5cix	B	1	127	1.3e-09	0.33	0.01		BACTERIOPHAGE P1 CRE GENE; CHAIN: A, B; DNA (35-MER); CHAIN: C, D;	PROTEIN/DNA CRE RECOMBINASE, DNA BENDING, SITE SPECIFIC RECOMBINATION, 2 PROTEIN-DNA INTERACTION, PROTEIN/DNA
1987	1awq	A	2	170	1.3e-52			127.69	CYCLOPHILIN A; CHAIN: A; PEPTIDE FROM THE HIV-1 CAPSID PROTEIN; CHAIN: B;	COMPLEX (ISOMERASE/PEPTIDE) COMPLEX (ISOMERASE/PEPTIDE), CYCLOPHILIN A, HIV-1 CAPSID,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1987	1cyn	A	3	182	1.1e-45			99.73	CYCLOPHILIN B; 1CYN 6 CHAIN: A; 1CYN 7 (D- (CHOLINYL)ALA)8- CYCLOSPORIN, 1CYN 10 CHAIN: C; 1CYN 11	2 PSEUDO-SYMMETRY COMPLEX (ISOMERASE/MMUNOSUPPRES SANT) CYCLOSPORIN, ISOMERASE, ROTAMASE, SIGNAL 1CYN 19
1987	2rmc	A	1	178	9.6e-45			97.55	COMPLEX (ISOMERASE/MMUNOSUPPR ESSANT) CYCLOPHILIN C COMPLEXED WITH CYCLOSPORIN A 2RMC 3	
1999	1ez3	A	85	143	2.4e-10	1.30	-0.14		SYNTAXIN-1A; CHAIN: A, B, C;	ENDOCYTOSIS/EXOCYTOSIS SYNAPTOTAGMIN ASSOCIATED 35 KDA PROTEIN, P35A, THREE HELIX BUNDLE
1999	1quu	A	86	143	1.7e-10	0.81	-0.19		HUMAN SKELETAL MUSCLE ALPHA-ACTININ 2; CHAIN: A;	CONTRACTILE PROTEIN TRIPLE-HELIX COILED COIL, CONTRACTILE PROTEIN
1999	1req	A	28	143	1.4e-09	0.11	-0.17		METHYLMALONYL-COA MUTASE; CHAIN: A, B, C, D;	ISOMERASE ISOMERASE, MUTASE, INTRAMOLECULAR TRANSFERASE
1999	1req	A	86	143	1e-10	0.55	-0.20		METHYLMALONYL-COA MUTASE; CHAIN: A, B, C, D;	ISOMERASE ISOMERASE, MUTASE, INTRAMOLECULAR TRANSFERASE
1999	1req	A	88	153	3.4e-09	0.02	-0.20		METHYLMALONYL-COA MUTASE; CHAIN: A, B, C, D;	ISOMERASE ISOMERASE, MUTASE, INTRAMOLECULAR TRANSFERASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
1999	2tc	P	86	143	2.7e-13	0.50	-0.19		TRANSDUCIN; CHAIN: B, G; PHOSDUCIN; CHAIN: P;	COMPLEX (TRANSDUCER/TRANSDUCTION ) GT BETA-GAMMA; MEKA, PP33; PHOSDUCIN, TRANSDUCIN, BETA-GAMMA, SIGNAL TRANSDUCTION, 2 REGULATION, PHOSPHORYLATION, G PROTEINS, THIOREDOXIN, 3 VISION, MEKA, COMPLEX (TRANSDUCER/TRANSDUCTION )
2001	1dxx	A	237	384	4.8e-37	-0.09	0.10		DYSTROPHIN; CHAIN: A, B, C, D;	STRUCTURAL PROTEIN DYSTROPHIN, MUSCULAR DYSTROPHY, CALPONTIN HOMOLOGY DOMAIN, 2 ACTIN-BINDING, UTROPHIN
2001	1qag	A	237	384	1.6e-35	-0.21	0.27		UTROPHIN ACTIN BINDING REGION; CHAIN: A, B;	STRUCTURAL PROTEIN CALPONTIN HOMOLOGY DOMAIN, DOMAIN SWAPPING, ACTIN BINDING, 2 UTROPHIN, DYSTROPHIN, STRUCTURAL PROTEIN
2011	1bhh	A	38	270	3.2e-28	0.01	-0.15		HEMOLIN; CHAIN: A, B;	INSECT IMMUNITY INSECT IMMUNITY, LPS-BINDING, HOMOPHILIC ADHESION



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2011	1cs6	A	38	271	4.8e-40	0.17	-0.18		AXONIN-1; CHAIN: A;	CELL ADHESION NEURAL CELL ADHESION
2011	1cvs	C	170	287	1.6e-16	0.07	0.69		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH FACTOR/GROWTH FACTOR RECEPTOR
2011	1cvs	C	42	161	1.6e-20	0.09	-0.20		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH FACTOR/GROWTH FACTOR RECEPTOR
2011	1cvs	C	77	270	1.4e-48	0.14	-0.02		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH FACTOR/GROWTH FACTOR RECEPTOR
2011	1cvs	D	170	287	1.6e-16	0.07	0.47		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Pst Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2011	levs	D	77	270	3.2e-45	0.03	-0.07		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	FACTOR/GROWTH FACTOR RECEPTOR
2011	1d5i	L	84	256	1.6e-13	0.14	-0.17		CHIMERIC GERMLINE PRECURSOR OF OXY-COPE CHAIN: L; CHIMERIC GERMLINE PRECURSOR OF OXY-COPE CHAIN: H;	IMMUNE SYSTEM IMMUNE SYSTEM
2011	1epf	A	83	254	3.2e-21	0.26	-0.15		NEURAL CELL ADHESION MOLECULE; CHAIN: A, B, C, D;	CELL ADHESION NCAM; NCAM, IMMUNOGLOBULIN FOLD, GLYCOPROTEIN
2011	lev2	E	79	270	3.2e-41	0.01	-0.11		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF2, FGFR2; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
2011	lev2	G	181	289	1.3e-15	-0.17	0.11		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF2, FGFR2; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									E, F, G, H;	TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
2011	lev2	G	79	274	4.8e-45	0.02	-0.09		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGFR2; FGFR2; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
2011	levt	C	170	287	3.2e-16	0.06	0.19		FIBROBLAST GROWTH FACTOR 1; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGFR1; FGFR1; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
2011	levt	C	42	161	8e-22	0.06	-0.20		FIBROBLAST GROWTH FACTOR 1; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGFR1; FGFR1; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
2011	lfbg	A	165	270	1.6e-20	0.55	0.93		TELOKIN; CHAIN: A	CONTRACTILE PROTEIN IMMUNOGLOBULIN FOLD, BETA BARREL
2011	lmcw	W	59	272	1.1e-08			51.74	IMMUNOGLOBULIN	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
2011	1nct		172	270	1e-19	0.41	0.25		IMMUNOGLOBULIN HETEROLOGOUS LIGHT CHAIN DIMER 1MCW 3 (MCGS-/WEIRS HYBRID) 1MCW 4	
									TTTN; CHAIN: NULL;	MUSCLE PROTEIN CONNECTIN, NEXTM5; CELL ADHESION, GLYCOPROTEIN, TRANSMEMBRANE, REPEAT, BRAIN, 2 IMMUNOGLOBULIN FOLD, ALTERNATIVE SPLICING, SIGNAL, 3 MUSCLE PROTEIN
2011	1lbr	R	51	170	1e-11			53.60	THROMBIN; CHAIN: L, H, J, K; RHODNIN; CHAIN: R, S;	COMPLEX (SERINE PROTEASE/INHIBITOR) COMPLEX (SERINE PROTEASE/INHIBITOR), KAZAL-TYPE INHIBITOR, 2 THROMBIN
2011	1ttn		173	271	6.4e-17	0.22	0.09		MUSCLE PROTEIN TITIN MODULE M5 (CONNECTIN) 1TNM 3 (NMR, MINIMIZED AVERAGE STRUCTURE) 1TNM 4 1TNM 58	
2011	1wit		172	270	3.4e-20	0.19	-0.06		TWITCHIN 18TH IGSF MODULE; CHAIN: NULL;	MUSCLE PROTEIN IMMUNOGLOBULIN SUPERFAMILY, 1 SET, MUSCLE PROTEIN
2011	1www	X	175	273	2.4e-21	0.23	-0.02		NERVE GROWTH FACTOR; CHAIN: V, W, TRKA	NERVE GROWTH FACTOR/TRKA COMPLEX

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
2011	2mcg	1	59	277	6.4e-08			55.90	IMMUNOGLOBULIN LAMBDA LIGHT CHAIN DIMER (MCG\$) 2MCG 3 (TRIGONAL FORM) 2MCG 4	BETA-NGF; COMPLEX, TRKA RECEPTOR, NERVE GROWTH FACTOR, CYSTEINE KNOT, 2 IMMUNOGLOBULIN LIKE DOMAIN, NERVE GROWTH FACTOR/TRKA COMPLEX
2011	9wga	A	26	186	0.00068			50.51	LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
2015	1a75	A	334	381	0.0001	-0.14	0.16		PARVALBUMIN; CHAIN: A, B	CALCIUM BINDING PROTEIN CALCIUM BINDING PROTEIN, MUSCLE PROTEIN
2015	1aj4		340	381	0.00014	0.22	0.83		TROPONIN C; CHAIN: NDL1;	MUSCLE PROTEIN CTNC; CARDIAC, MUSCLE PROTEIN, REGULATORY, CALCIUM BINDING
2015	1ak8		340	382	3.4e-06	-0.40	0.42		CALMODULIN; CHAIN: NDL1;	CALCIUM-BINDING PROTEIN CALMODULIN CERUM TRIC-DOMAIN, RESIDUES 1 - 75; CERUM-LOADED, CALCIUM-BINDING PROTEIN
2015	1ap4		340	381	0.00014	0.44	0.72		CARDIAC N-TROPONIN C;	CALCIUM-BINDING CTNC;

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN: NULL;	CALCIUM-BINDING, REGULATION, TROPONIN C, CARDIAC MUSCLE 2 CONTRACTION
2015	1br1	B	340	389	1.7e-06	0.24	0.74		MYOSIN; CHAIN: A, B, C, D, E, F, G, H;	MUSCLE PROTEIN MDE; MUSCLE PROTEIN
2015	1cdm	A	340	382	2.7e-06	0.34	0.74		CALCIUM-BINDING PROTEIN CALMODULIN COMPLEXED WITH CALMODULIN-BINDING DOMAIN OF ICDM 3 CALMODULIN-DEPENDENT PROTEIN KINASE II ICDM 4	
2015	1cll		340	382	2.4e-06	0.04	0.64		CALCIUM-BINDING PROTEIN CALMODULIN (VERTEBRATE) ICLL 3	
2015	1cmf		340	381	3.4e-06	0.46	0.76		CALMODULIN (VERTEBRATE); ICMF 6 CHAIN: NULL; ICMF 7	CALCIUM-BINDING PROTEIN CALMODULIN APO TR2C-DOMAIN; ICMF 9
2015	1dix	B	353	382	1.4e-06	0.14	0.22		PHOSPHOINOSITIDE-SPECIFIC PHOSPHOLIPASE C; CHAIN: A, B;	LIPID DEGRADATION PLC-D1; PHOSPHORIC DIESTER HYDROLASE, HYDROLASE, LIPID DEGRADATION, 2 TRANSDUCER, CALCIUM-BINDING, PHOSPHOLIPASE C, 3 PHOSPHOINOSITIDE-SPECIFIC STRUCTURAL PROTEIN HELIX-TURN-HELIX
2015	1dtl	A	340	382	0.00014	0.07	0.66		CARDIAC TROPONIN C; CHAIN: A;	OXIDOREDUCTASE FOUR EF-
2015	1ej3	A	338	381	3.1e-06	-0.07	0.33		AEQUORIN; CHAIN: A, B;	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2015	1f71	A	340	381	1.4e-06	0.20	0.92		CALMODULIN; CHAIN: A;	HAND CALCIUM-BINDING PROTEIN, PROTEIN-2 COELENTERAZINE PEROXIDE COMPLEX
2015	1fpw	A	338	381	1.7e-05	-0.47	0.00		CALCIUM-BINDING PROTEIN NCS-1; CHAIN: A;	TRANSPORT PROTEIN CALCIUM BINDING, EF HAND, FOUR-HELIX BUNDLE
2015	1mai		165	280	1.6e-34	1.11	1.00		PHOSPHOLIPASE C DELTA-1; CHAIN: NULL;	METAL BINDING PROTEIN YEAST FREQUENIN EF-HAND, CALCIUM
2015	1mai		165	281	1.6e-34			63.96	PHOSPHOLIPASE C DELTA-1; CHAIN: NULL;	SIGNAL TRANSDUCTION PROTEIN PLECKSTRIN, PHOSPHOLIPASE, INOSITOL TRISPHOSPHATE, 2 SIGNAL TRANSDUCTION PROTEIN, HYDROLASE
2015	1iro		340	382	0.0001	0.18	0.29		CALCIUM-BINDING PROTEIN RAT ONCOMODULIN IRRO 3	SIGNAL TRANSDUCTION PROTEIN PLECKSTRIN, PHOSPHOLIPASE, INOSITOL TRISPHOSPHATE, 2 SIGNAL TRANSDUCTION PROTEIN, HYDROLASE
2015	1vfk	A	340	382	6.8e-06	-0.27	0.88		CALMODULIN; CHAIN: A; RS20; CHAIN: B;	CALMODULIN, CALCIUM BINDING, HELIX-LOOP-HELIX, SIGNALING, 2 COMPLEX/CALCIUM-BINDING

Table 5										
SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQROL D score	Compound	PDB annotation
2015	1wdc	B	340	381	1e-05	-0.16	0.05		SCALLOP MYOSIN; CHAIN: A, B, C;	PROTEIN/PEPTIDE) MUSCLE PROTEIN MYOSIN, CALCIUM BINDING PROTEIN, MUSCLE PROTEIN
2015	1wdc	C	340	381	6.8e-06	0.06	0.54		SCALLOP MYOSIN; CHAIN: A, B, C;	MUSCLE PROTEIN MYOSIN, CALCIUM BINDING PROTEIN, MUSCLE PROTEIN
2025	1ady	A	44	351	1.3e-41	-0.00	0.04		RIBONUCLEASE INHIBITOR; CHAIN: A, D; ANGIOGENIN; CHAIN: B, E;	COMPLEX (INHIBITOR/NUCLEASE) COMPLEX (RI-ANG), HYDROLASE 2 MOLECULAR RECOGNITION, EPITOPE MAPPING, LEUCINE-RICH 3 REPEATS
2025	1fgv	A	60	341	3.2e-12	0.16	0.57		SKP2; CHAIN: A, C, E, G, I, K, M, O; SKP1; CHAIN: B, D, F, H, J, L, N, P;	LIGASE CYCLIN A/CDK2-ASSOCIATED PROTEIN P45; CYCLIN A/CDK2-ASSOCIATED BOX, LRR, LEUCINE-RICH REPEAT, SCF, UBIQUITIN, 2 E3, LIGASE CYCLIN A/CDK2-ASSOCIATED P45; CYCLIN A/CDK2-ASSOCIATED P19; SKP1, SKP2, F-REPEAT, SCF, UBIQUITIN, 2 E3, LIGASE CYCLIN A/CDK2-ASSOCIATED P45; CYCLIN A/CDK2-ASSOCIATED P19; SKP1, SKP2, F-BOX, LRRS, LEUCINE-
2025	1fs2	A	135	341	8e-10	0.48	0.60		SKP2; CHAIN: A, C; SKP1; CHAIN: B, D;	LIGASE CYCLIN A/CDK2-ASSOCIATED P45; CYCLIN A/CDK2-ASSOCIATED P19; SKP1, SKP2, F-BOX, LRRS, LEUCINE-



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2025	1yr8	A	147	344	3.2e-21	0.27	0.21		GTPASE-ACTIVATING PROTEIN RNA1_SCHPO; CHAIN: A, B;	RICH REPEATS, SCF, 2 UBIQUITIN, E3, UBIQUITIN PROTEIN LIGASE
2025	1yr8	A	67	322	1.4e-14	0.23	0.81		GTPASE-ACTIVATING PROTEIN RNA1_SCHPO; CHAIN: A, B;	TRANSCRIPTION RNAIP; RANGAP; GTPASE-ACTIVATING PROTEIN FOR SPI1, GTPASE-ACTIVATING PROTEIN, GAP, RNAIP, RANGAP, LRR, LEUCINE-2 RICH REPEAT PROTEIN, TWINNING, HEMIHEDRAL TWINNING, 3 MEROHEDRAL TWINNING, MEROHEDRY
2025	1yr8	A	70	275	6.4e-18	0.14	0.01		GTPASE-ACTIVATING PROTEIN RNA1_SCHPO; CHAIN: A, B;	TRANSCRIPTION RNAIP; RANGAP; GTPASE-ACTIVATING PROTEIN FOR SPI1, GTPASE-ACTIVATING PROTEIN, GAP, RNAIP, RANGAP, LRR,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2025	2bhh		45	351	3.2e-46	-0.13	0.05		RIBONUCLEASE INHIBITOR; CHAIN: NULL;	LEUCINE-2 RICH REPEAT PROTEIN, TWINNING, HEMIHEDRAL TWINNING, 3 MEROHEDRAL TWINNING, MEROHEDRY
										ACETYLATION RNASE INHIBITOR, RIBONUCLEASE/ANGIOGENIN INHIBITOR ACETYLATION, LEUCINE-RICH REPEATS
2034	1f21	A	123	257	1.6e-27	0.33	0.24		RIBONUCLEASE H; CHAIN: A;	HYDROLASE RNASE H, NUCLEASE, RNASE H*, RIBONUCLEASE H, METAL-BINDING 2 PROTEIN, PROTEIN FOLDING
2034	1hth	A	116	248	6.4e-22	0.15	0.59		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H DOMAIN OF /HIV-1S REVERSE TRANSCRIPTASE 1HRH 3	
2034	1rll		125	267	1.4e-23	0.27	0.96		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H (E.C.3.1.26.4) 1RLI 3	
2034	1rll		127	257	8e-22	0.30	0.29		HYDROLASE(ENDORIBONUCLEASE) RIBONUCLEASE H (E.C.3.1.26.4) 1RLI 3	
2034	1vrt	A	3	244	4.8e-45	-0.07	0.01		HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN: A, B, IVRT 5	REVERSE TRANSCRIPTASE IVRT 15
2052	1asu		156	311	1.6e-23			60.78	AVIAN SARCOMA VIRUS INTEGRASE; 1ASU 7 CHAIN: NULL; 1ASU 8	DNA INTEGRATION
2052	1asu		163	300	1.6e-23	0.39	0.78		AVIAN SARCOMA VIRUS INTEGRASE; 1ASU 7 CHAIN: NULL; 1ASU 8	DNA INTEGRATION
2052	1b9d	A	172	299	1.6e-22	0.46	0.86		INTEGRASE; CHAIN: A;	TRANSFERASE DNA INTEGRATION
2052	1b9f	A	172	299	4.8e-27	0.71	0.96		INTEGRASE; CHAIN: A;	TRANSFERASE DNA INTEGRATION
2052	1bl3	C	157	313	1.6e-29	0.33	0.35		INTEGRASE; CHAIN: A, B, C;	DNA INTEGRATION; TRANSFERASE INTEGRATION; AIDS, POLYPROTEIN, HYDROLASE, 2 ENDONUCLEASE, POLYNUCLEOTIDYL TRANSFERASE, DNA BINDING 3 (VIRAL)
2052	1c0m	A	160	326	8e-27	0.19	0.40		INTEGRASE; CHAIN: A, B, C, D;	TRANSFERASE INTEGRASE, ROUS SARCOMA VIRUS, HIV, X- RAY CRYSTALLOGRAPHY, 2 PROTEIN STRUCTURE, TRANSFERASE
2052	1cla	B	168	326	3.2e-24	0.21	0.84		RSV INTEGRASE; CHAIN: A, B;	VIRUS/VIRAL PROTEIN INTEGRASE, ROUS SARCOMA VIRUS, HIV, X-RAY

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2052	1exq	A	164	300	1.4e-20			51.79	AVIAN SARCOMA VIRUS INTEGRASE; CHAIN: A;	CRYSTALLOGRAPHY, 2 VIRUS/VIRAL PROTEIN
2052	1cz9	A	166	299	6.8e-23			58.62	AVIAN SARCOMA VIRUS INTEGRASE; CHAIN: A;	TRANSFERASE MIXED BETA-SHEET SURROUNDED BY ALPHA-HELICES
2052	1cz9	A	166	299	6.8e-23	0.39	1.00		AVIAN SARCOMA VIRUS INTEGRASE; CHAIN: A;	TRANSFERASE MIXED BETA-SHEET SURROUNDED BY ALPHA-HELICES
2052	1ex4	A	172	313	3.2e-24	0.52	0.88		INTEGRASE; CHAIN: A, B;	VIRUS/VIRAL PROTEIN SH3-LIKE DOMAIN, NONSPECIFIC DNA BINDING BETA SHEET, CIS-2 PROLINE
2052	1exq	A	172	313	8e-22	0.37	0.81		POLYPROTEIN; CHAIN: A, B;	VIRUS/VIRAL PROTEIN HIV-1 INTEGRASE, POLYNUCLEOTIDYL TRANSFERASE, DNA-BINDING 2 PROTEIN, DD35E
2052	1qs4	A	172	313	1.6e-24	0.47	0.94		HIV-1 INTEGRASE; CHAIN: A, B, C;	HYDROLASE DNA INTEGRATION, INTEGRASE, HIV, HYDROLASE, ASPARTYL 2 PROTEASE, ENDONUCLEASE
2058	1asu		2	134	1.1e-22	0.15	0.89		AVIAN SARCOMA VIRUS INTEGRASE; 1ASU 7 CHAIN: NULL; 1ASU 8	DNA INTEGRATION

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2058	1b9d	A	8	138	1.3e-27	-0.03	0.63		INTEGRASE; CHAIN: A;	TRANSFERASE DNA INTEGRATION
2058	1b9f	A	8	138	1.6e-33	0.04	0.69		INTEGRASE; CHAIN: A;	TRANSFERASE DNA INTEGRATION, TRASFERASE
2058	1b13	C	8	132	1.4e-35	0.13	0.98		INTEGRASE; CHAIN: A, B, C;	DNA INTEGRATION DNA INTEGRATION, AIDS, POLYPROTEIN, HYDROLASE, 2 ENDONUCLEASE, POLYNUCLEOTIDYL TRANSFERASE, DNA BINDING 3 (VIRAL)
2058	1c0m	A	1	145	3.4e-27	0.29	0.84		INTEGRASE; CHAIN: A, B, C, D;	TRANSFERASE INTEGRASE, ROUS SARCOMA VIRUS, HIV, X-RAY CRYSTALLOGRAPHY, 2 PROTEIN STRUCTURE, TRANSFERASE
2058	1c0m	A	2	157	1.6e-26	-0.09	0.72		INTEGRASE; CHAIN: A, B, C, D;	TRANSFERASE INTEGRASE, ROUS SARCOMA VIRUS, HIV, X-RAY CRYSTALLOGRAPHY, 2 PROTEIN STRUCTURE, TRANSFERASE
2058	1c1a	B	2	157	1.4e-24	0.24	0.59		RSV INTEGRASE; CHAIN: A, B;	VIRUS/VIRAL PROTEIN INTEGRASE, ROUS SARCOMA VIRUS, HIV, X-RAY CRYSTALLOGRAPHY, 2
2058	1ex4	A	8	132	9.6e-30	-0.15	0.92		INTEGRASE; CHAIN: A, B;	VIRUS/VIRAL PROTEIN SH3-LIKE DOMAIN, NONSPECIFIC

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2058	1exq	A	8	132	1.6e-27	-0.09	0.76		POL POLYPROTEIN; CHAIN: A, B;	DNA BINDING BETA SHEET, CIS-2 PROLINE
2058	1qst	A	8	132	3.2e-30	0.14	0.84		HIV-1 INTEGRASE; CHAIN: A, B, C;	HYDROLASE DNA INTEGRATION, INTEGRASE, HIV, HYDROLASE, ASPARTYL 2 PROTEASE, ENDONUCLEASE
2070	1gig	H	74	292	0.0045			52.34	IMMUNOGLOBULIN IGG1 FAB FRAGMENT (HC19) IGG 3	
2070	1osp	H	74	290	0.0059			57.86	FAB 184.1; CHAIN: L, H; OUTER SURFACE PROTEIN A; CHAIN: O;	COMPLEX (IMMUNOGLOBULIN/LIPOPROT EIN) OSPA; COMPLEX (IMMUNOGLOBULIN/LIPOPROT EIN), OUTER SURFACE 2 PROTEIN A COMPLEXED WITH FAB184.1, BORRELLIA BURGDOFFER 3 STRAIN B31
2070	25c8	H	74	291	0.0045			52.67	IGG 5C8; CHAIN: L, H;	CATALYTIC ANTIBODY CATALYTIC ANTIBODY, FAB, RING CLOSURE REACTION
2074	1a1h	A	11	96	4.8e-27	-0.03	0.88		QGSR ZINC FINGER PEPTIDE; CHAIN: A; DUPLEX	COMPLEX (ZINC FINGER/DNA) COMPLEX (ZINC FINGER/DNA),

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									OLIGONUCLEOTIDE BINDING SITE; CHAIN: B, C;	ZINC FINGER, DNA-BINDING PROTEIN
2074	1a1h	A	40	124	1.4e-32	0.44	1.00		QGSZ ZINC FINGER PEPTIDE; CHAIN: A; DUPLEX OLIGONUCLEOTIDE BINDING SITE; CHAIN: B, C;	COMPLEX (ZINC FINGER/DNA) COMPLEX (ZINC FINGER/DNA), ZINC FINGER, DNA-BINDING PROTEIN
2074	1a1h	A	40	124	1.6e-32	0.44	1.00		QGSZ ZINC FINGER PEPTIDE; CHAIN: A; DUPLEX OLIGONUCLEOTIDE BINDING SITE; CHAIN: B, C;	COMPLEX (ZINC FINGER/DNA) COMPLEX (ZINC FINGER/DNA), ZINC FINGER, DNA-BINDING PROTEIN
2074	1a1h	A	40	126	3.2e-33			84.04	QGSZ ZINC FINGER PEPTIDE; CHAIN: A; DUPLEX OLIGONUCLEOTIDE BINDING SITE; CHAIN: B, C;	COMPLEX (ZINC FINGER/DNA) COMPLEX (ZINC FINGER/DNA), ZINC FINGER, DNA-BINDING PROTEIN
2074	1a1h	A	70	152	3.2e-33	0.09	0.46		QGSZ ZINC FINGER PEPTIDE; CHAIN: A; DUPLEX OLIGONUCLEOTIDE BINDING SITE; CHAIN: B, C;	COMPLEX (ZINC FINGER/DNA) COMPLEX (ZINC FINGER/DNA), ZINC FINGER, DNA-BINDING PROTEIN
2074	1mev	C	10	96	1.6e-45	0.02	0.95		DNA; CHAIN: A, B, D, E; CONSENSUS ZINC FINGER PROTEIN; CHAIN: C, F, G;	COMPLEX (ZINC FINGER/DNA) ZINC FINGER, PROTEIN-DNA INTERACTION, PROTEIN DESIGN, 2 CRYSTAL STRUCTURE, COMPLEX (ZINC FINGER/DNA)
2074	1mev	C	39	124	6.4e-49	0.27	1.00		DNA; CHAIN: A, B, D, E; CONSENSUS ZINC FINGER PROTEIN; CHAIN: C, F, G;	COMPLEX (ZINC FINGER/DNA) ZINC FINGER, PROTEIN-DNA INTERACTION, PROTEIN DESIGN, 2 CRYSTAL

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2074	1mev	C	39	125	4.8e-49			91.42	DNA; CHAIN: A, B, D, E; CONSENSUS ZINC FINGER PROTEIN; CHAIN: C, F, G;	STRUCTURE, COMPLEX (ZINC FINGER/DNA) COMPLEX (ZINC FINGER/DNA) ZINC FINGER, PROTEIN-DNA INTERACTION, PROTEIN DESIGN, 2 CRYSTAL STRUCTURE, COMPLEX (ZINC FINGER/DNA)
2074	1mev	C	69	152	4.8e-49	0.41	0.57		DNA; CHAIN: A, B, D, E; CONSENSUS ZINC FINGER PROTEIN; CHAIN: C, F, G;	COMPLEX (ZINC FINGER/DNA) ZINC FINGER, PROTEIN-DNA INTERACTION, PROTEIN DESIGN, 2 CRYSTAL STRUCTURE, COMPLEX (ZINC FINGER/DNA)
2074	1tE3	A	11	96	3.2e-23	0.01	0.35		TRANSCRIPTION FACTOR IIIA; CHAIN: A; 5S RNA GENE; CHAIN: E, F;	COMPLEX (TRANSCRIPTION REGULATION/DNA) TFIIIA; 5S GENE; NMR, TFIIIA, PROTEIN, DNA, TRANSCRIPTION FACTOR, 5S RNA 2 GENE, DNA BINDING PROTEIN, ZINC FINGER, COMPLEX 3 (TRANSCRIPTION REGULATION/DNA)
2074	1tE3	A	39	128	3.2e-24			75.53	TRANSCRIPTION FACTOR IIIA; CHAIN: A; 5S RNA GENE; CHAIN: E, F;	COMPLEX (TRANSCRIPTION REGULATION/DNA) TFIIIA; 5S GENE; NMR, TFIIIA, PROTEIN, DNA, TRANSCRIPTION FACTOR, 5S RNA 2 GENE, DNA BINDING PROTEIN, ZINC FINGER,



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2074	1tf3	A	40	124	3.2e-24	0.36	0.96		TRANSCRIPTION FACTOR IIIA; CHAIN: A; 5S RNA GENE; CHAIN: E, F;	COMPLEX 3 (TRANSCRIPTION REGULATION/DNA)
2074	1tf6	A	13	161	1.6e-38	0.11	0.13		TFIIIA; CHAIN: A, D; 5S RIBOSOMAL RNA GENE; CHAIN: B, C, E, F;	COMPLEX (TRANSCRIPTION REGULATION/DNA) COMPLEX (TRANSCRIPTION REGULATION/DNA), RNA POLYMERASE III, 2 TRANSCRIPTION INITIATION, ZINC FINGER PROTEIN
2074	1tf6	A	2	105	3.2e-23	-0.09	0.13		TFIIIA; CHAIN: A, D; 5S RIBOSOMAL RNA GENE; CHAIN: B, C, E, F;	COMPLEX (TRANSCRIPTION REGULATION/DNA) COMPLEX (TRANSCRIPTION REGULATION/DNA), RNA POLYMERASE III, 2 TRANSCRIPTION INITIATION, ZINC FINGER PROTEIN
2074	1tf6	A	40	168	3.2e-32	-0.15	0.21		TFIIIA; CHAIN: A, D; 5S RIBOSOMAL RNA GENE; CHAIN: B, C, E, F;	COMPLEX (TRANSCRIPTION REGULATION/DNA) COMPLEX (TRANSCRIPTION REGULATION/DNA), RNA

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQFOLD score	Compound	PDB annotation
2074	1t66	A	9	173	1.6e-38			79.15	TFIIA; CHAIN: A, D, 5S RIBOSOMAL RNA GENE; CHAIN: B, C, E, F;	POLYMERASE III, 2 TRANSCRIPTION INITIATION, ZINC FINGER PROTEIN
2074	1ubd	C	1	96	1.6e-33	-0.09	0.57		YY1; CHAIN: C; ADENO-ASSOCIATED VIRUS P5 INITIATOR ELEMENT DNA; CHAIN: A, B;	COMPLEX (TRANSCRIPTION REGULATION/DNA) YING-YANG 1; TRANSCRIPTION INITIATION, INITIATOR ELEMENT, YY1, ZINC 2 FINGER PROTEIN, DNA-PROTEIN RECOGNITION, 3 COMPLEX (TRANSCRIPTION REGULATION/DNA)
2074	1ubd	C	15	124	1.3e-36	-0.06	1.00		YY1; CHAIN: C; ADENO-ASSOCIATED VIRUS P5 INITIATOR ELEMENT DNA; CHAIN: A, B;	COMPLEX (TRANSCRIPTION REGULATION/DNA) YING-YANG 1; TRANSCRIPTION INITIATION, INITIATOR ELEMENT, YY1, ZINC 2 FINGER PROTEIN, DNA-PROTEIN RECOGNITION, 3 COMPLEX (TRANSCRIPTION REGULATION/DNA)

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2074	1ubd	C	15	125	1.3e-36			84.50	YY1; CHAIN: C; ADENO-ASSOCIATED VIRUS P5 INITIATOR ELEMENT DNA; CHAIN: A, B;	COMPLEX (TRANSCRIPTION REGULATION/DNA) YING-YANG 1; TRANSCRIPTION INITIATION, INITIATOR ELEMENT, YY1, ZINC 2 FINGER PROTEIN, DNA-PROTEIN RECOGNITION, 3 COMPLEX (TRANSCRIPTION REGULATION/DNA)
2074	1ubd	C	44	152	8e-34	-0.13	0.93		YY1; CHAIN: C; ADENO-ASSOCIATED VIRUS P5 INITIATOR ELEMENT DNA; CHAIN: A, B;	COMPLEX (TRANSCRIPTION REGULATION/DNA) YING-YANG 1; TRANSCRIPTION INITIATION, INITIATOR ELEMENT, YY1, ZINC 2 FINGER PROTEIN, DNA-PROTEIN RECOGNITION, 3 COMPLEX (TRANSCRIPTION REGULATION/DNA)
2074	2adr		70	130	3.2e-16			56.22	ADRI; CHAIN: NULL;	TRANSCRIPTION REGULATION TRANSCRIPTION REGULATION, ADRI, ZINC FINGER, NMR
2074	2gli	A	11	156	3.2e-37			86.26	ZINC FINGER PROTEIN GLI1; CHAIN: A; DNA; CHAIN: C, D;	COMPLEX (DNA-BINDING PROTEIN/DNA) FIVE-FINGER GLI; GLI, ZINC FINGER, COMPLEX (DNA-BINDING PROTEIN/DNA)
2074	2gli	A	19	154	3.2e-37	0.11	0.28		ZINC FINGER PROTEIN GLI1; CHAIN: A; DNA; CHAIN: C, D;	COMPLEX (DNA-BINDING PROTEIN/DNA) FIVE-FINGER

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2074	2gi	A	4	123	3.2e-37	0.21	0.93		ZINC FINGER PROTEIN GLI1; CHAIN: A; DNA; CHAIN: C, D;	GLI, GLI, ZINC FINGER, COMPLEX (DNA-BINDING PROTEIN/DNA)
2074	2gi	A	44	181	4.8e-31	0.05	-0.05		ZINC FINGER PROTEIN GLI1; CHAIN: A; DNA; CHAIN: C, D;	COMPLEX (DNA-BINDING PROTEIN/DNA) FIVE-FINGER GLI, GLI, ZINC FINGER, COMPLEX (DNA-BINDING PROTEIN/DNA)
2076	1ad4	A	56	208	3.4e-15	0.08	0.45		RIBONUCLEASE INHIBITOR; CHAIN: A, D; ANGIOGENIN; CHAIN: B, E;	COMPLEX (INHIBITOR/NUCLEASE) COMPLEX (INHIBITOR/NUCLEASE), COMPLEX (RI-ANG), HYDROLASE 2 MOLECULAR RECOGNITION, EPTIPE MAPPING, LEUCINE-RICH 3 REPEATS
2076	1adn	A	40	109	9.6e-08	0.10	0.95		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A'; CHAIN: A, C; U2 B'; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP RIBONUCLEOPROTEIN
2076	1adn	A	59	145	3.4e-09	-0.04	0.47		U2 RNA HAIRPIN IV; CHAIN:	COMPLEX (NUCLEAR

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									Q, R; U2 A'; CHAIN: A, C; U2 B"; CHAIN: B, D;	PROTEIN/RNA COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
2076	1a9n	A	66	187	1.7e-22	0.46	1.00		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A'; CHAIN: A, C; U2 B"; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
2076	1a9n	C	40	109	9.6e-08	0.11	0.87		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A'; CHAIN: A, C; U2 B"; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
2076	1a9n	C	59	158	6.8e-10	0.18	0.63		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A'; CHAIN: A, C; U2 B"; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
2076	1a9n	C	66	187	6.8e-22	0.17	0.98		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A'; CHAIN: A, C; U2 B"; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
2076	1a9n	C	89	200	1.7e-16	0.43	0.16		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A'; CHAIN: A, C; U2 B"; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
2076	1d0b	A	36	199	1.6e-23	0.57	0.66		INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION
2076	1d0b	A							INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION
2076	1d0b	A	66	247	9.6e-24	0.41	0.92		INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
2076	1dce	A	38	116	3.2e-11	0.06	0.84		RAB GERANYLGERANYLTRANSFE RASE ALPHA SUBUNIT; CHAIN: A, C; RAB GERANYLGERANYLTRANSFE RASE BETA SUBUNIT; CHAIN: B, D;	CELL ADHESION TRANSFERASE CRYSTAL STRUCTURE, RAB GERANYLGERANYLTRANSFER ASE, 2.0 A 2 RESOLUTION, N- FORMYLMETHIONINE, ALPHA SUBUNIT, BETA SUBUNIT
2076	1dce	A	57	162	1.6e-12	0.13	0.89		RAB GERANYLGERANYLTRANSFE RASE ALPHA SUBUNIT; CHAIN: A, C; RAB GERANYLGERANYLTRANSFE RASE BETA SUBUNIT; CHAIN: B, D;	TRANSFERASE CRYSTAL STRUCTURE, RAB GERANYLGERANYLTRANSFER ASE, 2.0 A 2 RESOLUTION, N- FORMYLMETHIONINE, ALPHA SUBUNIT, BETA SUBUNIT
2076	1dce	A	77	208	1e-16	0.35	0.98		RAB GERANYLGERANYLTRANSFE RASE ALPHA SUBUNIT; CHAIN: A, C; RAB GERANYLGERANYLTRANSFE RASE BETA SUBUNIT; CHAIN: B, D;	TRANSFERASE CRYSTAL STRUCTURE, RAB GERANYLGERANYLTRANSFER ASE, 2.0 A 2 RESOLUTION, N- FORMYLMETHIONINE, ALPHA SUBUNIT, BETA SUBUNIT
2076	1ds9	A	59	163	3.2e-12	-0.38	0.31		OUTER ARM DYNEIN; CHAIN: A;	CONTRACTILE PROTEIN LEUCINE-RICH REPEAT, BETA- BETA-ALPHA CYLINDER, DYNEIN, 2 CHLAMYDOMONAS, FLAGELLA
2076	1yrg	A	37	163	4.8e-07	-0.05	0.24		GTPASE-ACTIVATING	TRANSCRIPTION RNAIP;

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									PROTEIN RNAI_SCHPO; CHAIN: A, B;	RANGAP; GTPASE-ACTIVATING PROTEIN FOR SPI1, GTPASE- ACTIVATING PROTEIN, GAP, RNAIP, RANGAP, LRR, LEUCINE-2 RICH REPEAT PROTEIN, TWINNING, HEMIHEDRAL TWINNING, 3 MEROHEDRAL TWINNING, MEROHEDRY
2076	1yr2	A	59	186	6.8e-14	-0.06	0.63		GTPASE-ACTIVATING PROTEIN RNAI_SCHPO; CHAIN: A, B;	TRANSCRIPTION RNAIP; RANGAP; GTPASE-ACTIVATING PROTEIN FOR SPI1, GTPASE- ACTIVATING PROTEIN, GAP, RNAIP, RANGAP, LRR, LEUCINE-2 RICH REPEAT PROTEIN, TWINNING, HEMIHEDRAL TWINNING, 3 MEROHEDRAL TWINNING, MEROHEDRY

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1042	28	0.969	0.829
1043	19	0.891	0.574
1044	26	0.953	0.774
1045	13	0.891	0.675
1046	19	0.987	0.941
1047	24	0.969	0.817
1048	11	0.953	0.814
1049	17	0.923	0.602
1050	26	0.977	0.685
1051	39	0.978	0.765
1052	22	0.982	0.918
1053	15	0.989	0.965
1054	24	0.912	0.655
1055	31	0.885	0.603
1056	27	0.924	0.593
1057	14	0.907	0.696
1058	22	0.945	0.759
1059	29	0.917	0.690
1060	21	0.973	0.669
1061	19	0.891	0.574
1062	16	0.924	0.790
1063	16	0.951	0.883
1064	23	0.913	0.702
1065	27	0.948	0.670
1066	17	0.903	0.714
1067	20	0.923	0.683
1068	18	0.987	0.939
1069	16	0.969	0.904
1070	19	0.991	0.955
1071	31	0.969	0.810
1072	17	0.926	0.683
1073	22	0.956	0.916
1074	20	0.989	0.903
1075	15	0.899	0.790
1076	15	0.990	0.963
1077	25	0.901	0.586
1078	13	0.908	0.661
1079	20	0.901	0.669
1080	17	0.963	0.692
1081	13	0.891	0.675
1082	20	0.944	0.831
1083	17	0.961	0.880
1084	34	0.888	0.611
1085	26	0.920	0.700
1086	21	0.948	0.853
1087	28	0.963	0.728
1088	22	0.987	0.828
1089	22	0.979	0.946
1090	26	0.908	0.557
1091	27	0.978	0.831
1092	13	0.971	0.905
1093	19	0.939	0.711
1094	35	0.938	0.657
1095	16	0.909	0.828
1096	18	0.937	0.773



Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1097	21	0.994	0.969
1098	15	0.949	0.849
1099	27	0.903	0.644
1100	21	0.987	0.895
1101	31	0.923	0.626
1102	25	0.986	0.932
1103	33	0.998	0.887
1104	23	0.990	0.932
1105	19	0.936	0.685
1106	27	0.910	0.566
1107	24	0.915	0.567
1108	15	0.937	0.732
1109	21	0.950	0.801
1110	25	0.965	0.890
1111	11	0.953	0.814
1112	33	0.963	0.577
1113	20	0.935	0.834
1114	14	0.938	0.795
1115	32	0.942	0.655
1116	23	0.957	0.596
1117	19	0.886	0.594
1118	23	0.994	0.966
1119	26	0.939	0.810
1120	18	0.930	0.656
1121	22	0.967	0.697
1122	18	0.983	0.961
1123	18	0.896	0.737
1124	31	0.932	0.598
1125	23	0.989	0.959
1126	18	0.960	0.753
1127	23	0.965	0.785
1128	33	0.969	0.791
1129	48	0.987	0.614
1130	15	0.975	0.934
1131	20	0.986	0.933
1132	22	0.981	0.883
1133	24	0.941	0.732
1134	18	0.916	0.728
1135	18	0.926	0.701
1136	31	0.971	0.816
1137	33	0.937	0.599
1138	27	0.922	0.559
1139	17	0.948	0.609
1140	24	0.985	0.945
1141	19	0.881	0.618
1142	27	0.932	0.726
1143	24	0.977	0.812
1144	25	0.948	0.848
1145	19	0.973	0.819
1146	20	0.955	0.612
1147	28	0.974	0.846
1148	14	0.944	0.864
1149	40	0.993	0.932
1150	16	0.969	0.912
1151	25	0.927	0.727

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1152	22	0.939	0.684
1153	32	0.925	0.578
1154	21	0.962	0.823
1155	19	0.944	0.719
1156	14	0.897	0.638
1159	31	0.982	0.594
1160	29	0.880	0.645
1161	19	0.970	0.823
1162	23	0.886	0.627
1163	22	0.983	0.953
1164	18	0.975	0.858
1166	29	0.924	0.661
1167	31	0.953	0.687
1168	23	0.967	0.832
1169	18	0.928	0.698
1170	18	0.968	0.806
1171	21	0.932	0.654
1172	20	0.932	0.660
1173	18	0.952	0.791
1174	16	0.900	0.629
1175	21	0.892	0.786
1176	27	0.979	0.837
1177	23	0.961	0.663
1178	23	0.974	0.782
1179	40	0.921	0.764
1180	25	0.966	0.910
1181	30	0.927	0.676
1183	22	0.942	0.807
1184	22	0.971	0.887
1185	33	0.963	0.851
1187	16	0.993	0.954
1188	17	0.940	0.789
1189	18	0.925	0.784
1190	18	0.965	0.733
1191	23	0.956	0.636
1192	31	0.992	0.803
1193	25	0.991	0.948
1194	20	0.927	0.617
1195	26	0.986	0.895
1196	30	0.889	0.618
1197	23	0.983	0.873
1198	30	0.993	0.815
1199	18	0.985	0.956
1201	6	0.885	0.564
1202	28	0.959	0.730
1203	29	0.916	0.707
1204	22	0.940	0.800
1205	16	0.888	0.646
1206	21	0.908	0.558
1207	27	0.953	0.564
1208	43	0.969	0.757
1209	27	0.965	0.891
1212	19	0.976	0.809
1213	20	0.988	0.872
1214	31	0.987	0.871

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1215	18	0.989	0.880
1216	34	0.920	0.550
1218	20	0.957	0.870
1219	25	0.928	0.615
1220	18	0.989	0.955
1221	14	0.892	0.686
1222	21	0.979	0.940
1223	24	0.979	0.930
1224	42	0.983	0.771
1225	22	0.982	0.811
1226	21	0.945	0.794
1227	15	0.969	0.910
1229	16	0.916	0.622
1230	29	0.972	0.769
1232	14	0.945	0.836
1233	30	0.963	0.669
1234	29	0.989	0.867
1235	34	0.977	0.891
1236	36	0.934	0.673
1237	32	0.922	0.720
1238	22	0.950	0.828
1239	22	0.956	0.763
1240	24	0.981	0.938
1241	19	0.891	0.574
1242	32	0.974	0.869
1243	33	0.890	0.675
1244	25	0.934	0.593
1245	22	0.944	0.709
1246	39	0.940	0.714
1247	29	0.889	0.658
1248	19	0.883	0.749
1249	24	0.892	0.577
1250	21	0.916	0.662
1251	29	0.921	0.601
1252	17	0.954	0.741
1253	27	0.888	0.738
1254	28	0.983	0.920
1256	26	0.975	0.705
1257	19	0.914	0.698
1258	18	0.961	0.869
1259	41	0.962	0.600
1260	18	0.947	0.664
1261	18	0.946	0.739
1262	20	0.889	0.561
1263	31	0.973	0.865
1264	18	0.956	0.850
1265	14	0.952	0.875
1266	29	0.902	0.563
1267	20	0.966	0.739
1268	23	0.953	0.688
1269	38	0.919	0.676
1270	27	0.955	0.826
1271	23	0.913	0.702
1273	21	0.972	0.915
1274	23	0.950	0.578

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1275	20	0.996	0.965
1276	20	0.976	0.937
1278	26	0.962	0.752
1279	38	0.962	0.756
1280	19	0.991	0.929
1281	27	0.948	0.670
1282	22	0.932	0.790
1283	23	0.962	0.679
1285	30	0.888	0.573
1286	15	0.996	0.988
1287	27	0.992	0.893
1288	24	0.952	0.685
1289	36	0.953	0.605
1290	32	0.932	0.649
1291	24	0.990	0.935
1292	24	0.973	0.940
1293	20	0.965	0.811
1294	18	0.977	0.957
1296	24	0.987	0.903
1297	12	0.894	0.780
1298	29	0.899	0.623
1299	19	0.882	0.753
1300	33	0.996	0.905
1301	21	0.952	0.663
1302	19	0.984	0.937
1303	32	0.978	0.885
1305	18	0.985	0.736
1306	46	0.991	0.888
1308	27	0.996	0.933
1309	24	0.970	0.913
1310	27	0.930	0.778
1312	16	0.990	0.959
1313	18	0.949	0.767
1314	18	0.896	0.752
1315	18	0.984	0.888
1316	21	0.953	0.721
1317	35	0.923	0.688
1318	27	0.940	0.796
1319	26	0.990	0.837
1320	24	0.972	0.663
1321	18	0.969	0.722
1323	21	0.955	0.709
1324	21	0.979	0.935
1325	26	0.944	0.675
1326	29	0.931	0.569
1327	18	0.997	0.955
1329	24	0.985	0.845
1330	43	0.901	0.602
1331	32	0.965	0.699
1332	15	0.881	0.608
1334	32	0.896	0.556
1335	18	0.963	0.807
1336	19	0.909	0.593
1337	16	0.885	0.562
1338	18	0.911	0.688

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1339	24	0.980	0.847
1340	25	0.943	0.774
1341	20	0.973	0.778
1342	27	0.924	0.686
1343	24	0.914	0.585
1344	16	0.957	0.773
1345	15	0.906	0.798
1346	16	0.971	0.855
1347	24	0.980	0.901
1348	23	0.965	0.642
1349	22	0.899	0.609
1350	18	0.940	0.585
1351	19	0.985	0.935
1352	22	0.945	0.718
1353	20	0.943	0.728
1354	15	0.887	0.721
1355	16	0.915	0.737
1358	21	0.948	0.585
1360	30	0.911	0.555
1361	20	0.976	0.851
1362	19	0.927	0.791
1364	19	0.947	0.574
1365	28	0.997	0.786
1366	28	0.979	0.855
1367	22	0.895	0.577
1368	19	0.956	0.829
1369	16	0.929	0.739
1370	17	0.931	0.745
1371	30	0.950	0.708
1372	28	0.968	0.856
1373	26	0.953	0.711
1375	32	0.983	0.842
1376	19	0.929	0.689
1377	30	0.899	0.631
1378	25	0.927	0.775
1379	19	0.982	0.922
1380	28	0.940	0.628
1381	20	0.890	0.610
1382	28	0.921	0.606
1383	23	0.881	0.644
1384	24	0.978	0.911
1385	21	0.974	0.723
1386	26	0.980	0.795
1387	16	0.903	0.654
1388	20	0.912	0.596
1389	19	0.981	0.960
1390	25	0.932	0.790
1391	15	0.990	0.963
1395	18	0.942	0.709
1396	28	0.963	0.844
1397	19	0.972	0.882
1398	21	0.966	0.827
1399	21	0.962	0.752
1400	25	0.979	0.855
1402	23	0.913	0.685

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1403	19	0.935	0.829
1404	21	0.984	0.958
1405	27	0.888	0.566
1406	36	0.945	0.564
1407	19	0.938	0.755
1408	22	0.947	0.745
1409	16	0.909	0.728
1410	20	0.961	0.866
1412	22	0.991	0.926
1413	20	0.911	0.683
1414	15	0.905	0.737
1416	13	0.933	0.799
1417	46	0.956	0.728
1418	20	0.945	0.782
1419	19	0.987	0.953
1420	30	0.976	0.862
1421	24	0.964	0.796
1423	23	0.924	0.645
1425	19	0.913	0.670
1426	33	0.968	0.774
1427	22	0.941	0.632
1428	18	0.972	0.935
1429	15	0.978	0.909
1430	26	0.926	0.713
1431	26	0.915	0.659
1432	21	0.949	0.790
1433	27	0.996	0.854
1434	26	0.910	0.590
1436	21	0.983	0.793
1437	18	0.932	0.643
1438	21	0.908	0.583
1439	24	0.925	0.742
1440	18	0.909	0.736
1441	30	0.883	0.615
1442	37	0.960	0.714
1444	30	0.942	0.586
1445	24	0.904	0.640
1446	26	0.950	0.724
1447	15	0.956	0.757
1448	30	0.906	0.692
1449	21	0.933	0.751
1450	25	0.990	0.855
1451	20	0.893	0.775
1452	26	0.952	0.729
1453	44	0.990	0.654
1454	20	0.974	0.810
1455	21	0.960	0.679
1456	17	0.926	0.629
1457	23	0.982	0.940
1458	18	0.986	0.938
1459	22	0.940	0.617
1460	18	0.939	0.698
1461	39	0.997	0.955
1462	11	0.989	0.626
1463	16	0.972	0.911

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1465	17	0.948	0.855
1466	13	0.901	0.739
1467	20	0.960	0.883
1468	26	0.903	0.585
1469	18	0.914	0.710
1470	23	0.972	0.908
1471	19	0.942	0.626
1473	25	0.972	0.670
1474	15	0.917	0.810
1475	40	0.923	0.825
1477	21	0.914	0.589
1478	26	0.964	0.721
1479	19	0.936	0.624
1481	22	0.995	0.943
1482	20	0.995	0.959
1484	19	0.964	0.755
1485	15	0.956	0.847
1486	27	0.963	0.584
1487	23	0.941	0.781
1488	32	0.969	0.816
1489	29	0.956	0.742
1491	20	0.894	0.615
1492	34	0.923	0.668
1493	16	0.943	0.809
1494	19	0.969	0.878
1495	27	0.944	0.726
1496	45	0.915	0.688
1497	45	0.908	0.583
1499	45	0.987	0.820
1500	20	0.972	0.790
1501	14	0.881	0.637
1503	24	0.973	0.786
1504	16	0.923	0.752
1505	22	0.965	0.829
1507	43	0.996	0.907
1509	21	0.948	0.732
1510	23	0.962	0.822
1511	34	0.921	0.646
1512	19	0.959	0.753
1513	46	0.962	0.628
1514	21	0.928	0.717
1515	16	0.926	0.731
1516	15	0.885	0.663
1517	21	0.935	0.795
1518	21	0.945	0.852
1519	13	0.881	0.636
1520	20	0.949	0.704
1521	21	0.938	0.745
1522	20	0.977	0.923
1523	23	0.925	0.619
1524	20	0.933	0.728
1525	11	0.912	0.784
1526	29	0.907	0.656
1527	18	0.962	0.704
1528	42	0.977	0.817

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1529	37	0.960	0.623
1530	22	0.899	0.649
1532	22	0.943	0.663
1533	20	0.970	0.936
1534	28	0.934	0.607
1535	30	0.989	0.890
1536	16	0.984	0.932
1537	22	0.992	0.974
1538	35	0.976	0.622
1539	20	0.901	0.576
1540	28	0.944	0.697
1542	28	0.936	0.667
1543	25	0.891	0.550
1544	21	0.967	0.700
1545	31	0.938	0.649
1546	21	0.883	0.569
1547	29	0.953	0.614
1548	12	0.916	0.815
1549	23	0.955	0.658
1550	21	0.948	0.635
1551	19	0.956	0.835
1552	18	0.960	0.803
1554	33	0.920	0.577
1555	24	0.947	0.717
1556	31	0.898	0.658
1557	24	0.960	0.876
1558	23	0.985	0.878
1560	38	0.919	0.553
1561	12	0.942	0.841
1562	21	0.887	0.568
1563	19	0.990	0.928
1564	18	0.950	0.814
1567	26	0.970	0.822
1569	14	0.928	0.806
1570	26	0.998	0.969
1571	18	0.911	0.762
1572	28	0.986	0.924
1574	15	0.935	0.815
1575	18	0.955	0.896
1576	26	0.949	0.697
1577	20	0.945	0.856
1578	24	0.962	0.723
1579	23	0.976	0.716
1580	20	0.903	0.597
1582	19	0.880	0.679
1583	25	0.984	0.918
1584	22	0.991	0.876
1585	23	0.968	0.710
1586	33	0.894	0.596
1587	23	0.918	0.721
1588	19	0.913	0.703
1589	14	0.951	0.886
1590	28	0.887	0.557
1591	26	0.999	0.969
1592	19	0.968	0.865



Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1593	32	0.962	0.612
1594	22	0.966	0.864
1596	19	0.970	0.823
1597	15	0.917	0.825
1598	32	0.991	0.900
1599	26	0.927	0.693
1600	18	0.896	0.656
1601	16	0.926	0.833
1602	18	0.948	0.883
1603	18	0.977	0.868
1604	34	0.943	0.730
1606	15	0.930	0.640
1607	32	0.967	0.697
1608	21	0.922	0.658
1610	30	0.881	0.586
1611	30	0.887	0.667
1612	19	0.938	0.565
1613	22	0.977	0.894
1614	20	0.925	0.725
1615	25	0.972	0.746
1616	30	0.986	0.671
1619	18	0.917	0.620
1620	28	0.968	0.611
1621	29	0.925	0.613
1622	48	0.968	0.711
1623	24	0.937	0.586
1624	19	0.914	0.694
1625	26	0.906	0.685
1626	14	0.962	0.863
1627	28	0.976	0.911
1629	17	0.973	0.938
1630	22	0.962	0.919
1632	31	0.997	0.846
1633	25	0.920	0.607
1634	17	0.982	0.945
1635	17	0.994	0.968
1638	30	0.922	0.705
1639	21	0.952	0.714
1640	21	0.966	0.807
1641	23	0.983	0.821
1642	18	0.953	0.885
1643	16	0.907	0.647
1644	20	0.884	0.650
1645	17	0.959	0.680
1646	18	0.991	0.954
1647	30	0.983	0.786
1648	21	0.886	0.567
1649	24	0.894	0.658
1650	23	0.881	0.657
1651	27	0.932	0.702
1652	22	0.993	0.885
1653	17	0.990	0.926
1654	19	0.932	0.622
1655	34	0.931	0.673
1656	19	0.966	0.909

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1657	17	0.955	0.867
1658	38	0.954	0.594
1659	19	0.920	0.710
1660	37	0.988	0.598
1662	32	0.909	0.675
1664	16	0.937	0.804
1665	20	0.911	0.621
1667	29	0.981	0.871
1668	33	0.972	0.869
1669	22	0.968	0.913
1670	23	0.990	0.932
1672	22	0.939	0.716
1673	17	0.963	0.865
1674	38	0.949	0.669
1675	20	0.926	0.787
1677	19	0.938	0.785
1678	20	0.929	0.727
1679	20	0.916	0.604
1680	21	0.967	0.886
1681	20	0.909	0.749
1682	30	0.928	0.776
1683	20	0.916	0.649
1684	21	0.976	0.879
1685	13	0.897	0.645
1686	13	0.994	0.963
1687	17	0.898	0.743
1688	30	0.946	0.638
1689	21	0.996	0.976
1690	18	0.916	0.595
1691	17	0.934	0.754
1692	28	0.899	0.753
1693	20	0.933	0.655
1694	19	0.990	0.920
1695	17	0.945	0.731
1697	18	0.885	0.588
1698	29	0.986	0.937
1699	26	0.972	0.557
1700	17	0.977	0.946
1701	17	0.882	0.608
1702	20	0.989	0.952
1703	22	0.919	0.578
1706	31	0.895	0.648
1707	22	0.965	0.922
1708	22	0.937	0.569
1709	20	0.980	0.903
1710	17	0.972	0.857
1711	27	0.984	0.823
1712	17	0.963	0.872
1713	24	0.977	0.880
1714	17	0.970	0.908
1715	31	0.973	0.843
1716	18	0.931	0.703
1717	18	0.931	0.702
1718	34	0.946	0.628
1719	19	0.973	0.883

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1720	48	0.980	0.845
1721	28	0.922	0.676
1722	44	0.965	0.645
1723	26	0.887	0.730
1724	25	0.939	0.795
1725	15	0.971	0.942
1727	23	0.923	0.591
1728	23	0.987	0.936
1729	18	0.927	0.814
1730	18	0.935	0.605
1731	25	0.972	0.912
1732	42	0.972	0.726
1733	20	0.952	0.798
1734	17	0.975	0.918
1735	15	0.979	0.877
1736	41	0.933	0.659
1738	17	0.925	0.746
1739	18	0.912	0.764
1741	11	0.953	0.814
1742	23	0.976	0.774
1744	23	0.918	0.606
1746	29	0.915	0.652
1747	15	0.933	0.840
1748	27	0.903	0.612
1750	29	0.904	0.618
1751	22	0.888	0.670
1752	16	0.979	0.868
1753	26	0.959	0.884
1754	22	0.954	0.696
1755	20	0.895	0.707
1756	26	0.906	0.703
1757	14	0.888	0.587
1758	15	0.994	0.953
1759	21	0.922	0.610
1760	21	0.942	0.693
1761	19	0.947	0.814
1762	21	0.934	0.655
1763	22	0.940	0.609
1764	23	0.937	0.832
1765	23	0.896	0.677
1766	26	0.909	0.690
1768	18	0.915	0.689
1769	36	0.969	0.602
1770	20	0.880	0.640
1772	20	0.942	0.715
1773	20	0.947	0.817
1774	16	0.969	0.880
1775	18	0.971	0.859
1776	24	0.891	0.670
1777	27	0.961	0.747
1778	40	0.963	0.574
1779	23	0.974	0.656
1780	21	0.899	0.653
1781	25	0.908	0.601
1782	19	0.943	0.678

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1783	23	0.936	0.634
1784	29	0.949	0.786
1785	44	0.915	0.571
1786	22	0.965	0.885
1787	15	0.974	0.940
1789	23	0.952	0.659
1790	16	0.972	0.898
1791	21	0.980	0.953
1792	32	0.961	0.668
1793	29	0.907	0.551
1794	22	0.957	0.934
1795	21	0.990	0.849
1796	22	0.954	0.893
1797	16	0.942	0.657
1799	25	0.949	0.840
1800	28	0.949	0.739
1801	25	0.938	0.767
1802	15	0.899	0.672
1803	17	0.987	0.956
1804	24	0.941	0.775
1805	26	0.972	0.771
1806	20	0.985	0.957
1807	22	0.932	0.571
1808	16	0.927	0.608
1809	26	0.987	0.770
1810	37	0.955	0.592
1811	28	0.911	0.632
1812	24	0.894	0.698
1813	22	0.906	0.624
1814	34	0.951	0.806
1816	25	0.919	0.578
1817	26	0.980	0.932
1818	19	0.993	0.940
1820	26	0.939	0.810
1821	48	0.967	0.556
1822	19	0.931	0.753
1823	36	0.892	0.670
1824	18	0.903	0.674
1825	17	0.966	0.854
1826	15	0.938	0.849
1827	27	0.985	0.891
1828	17	0.895	0.665
1829	36	0.916	0.620
1830	22	0.952	0.835
1831	17	0.961	0.731
1832	19	0.996	0.982
1833	19	0.918	0.556
1834	37	0.926	0.587
1836	14	0.897	0.787
1837	19	0.960	0.816
1838	31	0.902	0.632
1839	17	0.987	0.955
1840	23	0.988	0.941
1842	26	0.915	0.695
1843	26	0.987	0.926

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1844	15	0.933	0.731
1845	16	0.942	0.750
1846	20	0.914	0.842
1847	18	0.899	0.695
1848	24	0.988	0.883
1849	26	0.956	0.612
1850	31	0.961	0.568
1851	22	0.966	0.882
1853	30	0.921	0.610
1854	24	0.973	0.922
1855	14	0.938	0.902
1856	19	0.931	0.745
1857	21	0.908	0.556
1858	20	0.933	0.837
1859	23	0.920	0.633
1860	18	0.896	0.737
1862	16	0.887	0.641
1863	21	0.974	0.937
1864	24	0.982	0.899
1865	37	0.997	0.901
1867	19	0.960	0.758
1868	37	0.970	0.851
1869	20	0.950	0.684
1870	18	0.952	0.694
1871	16	0.921	0.724
1872	16	0.908	0.579
1873	24	0.991	0.913
1874	33	0.898	0.689
1875	26	0.904	0.707
1876	20	0.983	0.967
1877	18	0.951	0.739
1878	27	0.971	0.862
1879	45	0.966	0.761
1880	16	0.940	0.778
1881	35	0.926	0.704
1882	23	0.882	0.567
1883	19	0.933	0.703
1884	26	0.919	0.754
1886	25	0.911	0.570
1887	21	0.987	0.931
1888	39	0.965	0.616
1889	20	0.967	0.885
1890	23	0.980	0.871
1891	26	0.896	0.665
1892	20	0.882	0.729
1894	16	0.914	0.741
1895	28	0.997	0.888
1896	19	0.899	0.777
1897	17	0.893	0.615
1898	19	0.976	0.821
1899	22	0.952	0.791
1900	26	0.990	0.775
1901	16	0.985	0.958
1902	38	0.912	0.654
1903	26	0.952	0.870

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1904	25	0.949	0.844
1905	23	0.945	0.718
1906	18	0.907	0.556
1907	20	0.961	0.786
1908	19	0.907	0.752
1909	17	0.957	0.808
1910	22	0.933	0.778
1911	22	0.988	0.913
1912	32	0.964	0.814
1913	21	0.952	0.784
1914	24	0.946	0.644
1915	21	0.919	0.644
1916	21	0.969	0.912
1917	16	0.962	0.681
1918	14	0.926	0.776
1919	23	0.987	0.897
1920	48	0.987	0.614
1921	23	0.899	0.677
1922	23	0.907	0.651
1923	16	0.921	0.706
1924	20	0.928	0.672
1925	26	0.985	0.942
1926	27	0.911	0.682
1927	19	0.939	0.700
1928	15	0.887	0.709
1929	15	0.980	0.959
1930	25	0.987	0.924
1931	28	0.936	0.745
1932	20	0.958	0.669
1933	21	0.988	0.945
1934	24	0.912	0.699
1935	23	0.909	0.726
1936	20	0.964	0.924
1937	28	0.960	0.813
1938	18	0.971	0.806
1939	20	0.954	0.746
1941	20	0.986	0.933
1942	45	0.976	0.736
1944	18	0.967	0.871
1945	20	0.973	0.759
1947	17	0.954	0.919
1948	21	0.970	0.871
1949	18	0.991	0.976
1950	27	0.893	0.647
1951	19	0.881	0.705
1952	24	0.977	0.830
1953	15	0.957	0.834
1954	29	0.970	0.863
1956	19	0.940	0.835
1957	32	0.992	0.891
1958	22	0.968	0.837
1959	27	0.908	0.725
1960	20	0.941	0.751
1961	21	0.885	0.669
1962	29	0.955	0.797

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
1963	16	0.974	0.950
1964	21	0.929	0.745
1965	24	0.913	0.658
1966	45	0.937	0.671
1968	43	0.956	0.581
1969	19	0.956	0.614
1970	46	0.901	0.566
1971	24	0.947	0.768
1972	24	0.900	0.642
1974	22	0.988	0.922
1975	24	0.951	0.710
1976	18	0.932	0.740
1977	18	0.954	0.736
1978	20	0.994	0.967
1979	26	0.987	0.926
1980	22	0.964	0.866
1981	13	0.932	0.870
1982	21	0.949	0.881
1983	23	0.957	0.658
1984	12	0.954	0.910
1985	22	0.990	0.829
1986	31	0.987	0.845
1987	20	0.919	0.721
1988	17	0.985	0.966
1989	24	0.966	0.830
1990	31	0.971	0.816
1991	15	0.935	0.823
1992	21	0.967	0.802
1994	18	0.930	0.650
1995	20	0.902	0.611
1996	23	0.946	0.724
1997	25	0.943	0.787
1998	18	0.921	0.666
1999	13	0.883	0.748
2000	24	0.899	0.579
2001	13	0.918	0.705
2002	18	0.899	0.809
2003	18	0.950	0.647
2004	30	0.981	0.889
2005	17	0.950	0.771
2007	24	0.940	0.800
2008	21	0.980	0.815
2009	43	0.939	0.655
2010	16	0.920	0.698
2011	30	0.978	0.901
2012	19	0.981	0.919
2013	40	0.978	0.553
2014	20	0.994	0.960
2015	18	0.955	0.771
2016	25	0.914	0.769
2017	31	0.952	0.776
2018	26	0.985	0.854
2019	16	0.945	0.822
2020	22	0.973	0.804
2021	17	0.954	0.919

Table 6

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
2022	19	0.993	0.973
2023	18	0.921	0.683
2026	23	0.890	0.604
2027	35	0.943	0.603
2028	25	0.992	0.953
2029	47	0.950	0.846
2030	17	0.914	0.722
2032	18	0.995	0.974
2033	17	0.933	0.828
2034	17	0.934	0.644
2035	26	0.910	0.567
2036	30	0.940	0.690
2037	23	0.908	0.557
2038	18	0.906	0.624
2039	18	0.926	0.768
2040	14	0.934	0.758
2041	18	0.960	0.869
2042	21	0.911	0.716
2043	25	0.896	0.576
2044	27	0.953	0.850
2045	17	0.962	0.863
2046	25	0.924	0.572
2047	39	0.955	0.608
2048	38	0.958	0.692
2049	25	0.949	0.803
2050	27	0.932	0.726
2051	15	0.900	0.672
2052	22	0.967	0.703
2053	19	0.960	0.757
2054	20	0.880	0.775
2055	19	0.913	0.721
2057	23	0.955	0.882
2058	23	0.893	0.728
2059	26	0.953	0.619
2060	19	0.935	0.770
2061	44	0.952	0.739
2062	31	0.964	0.894
2063	19	0.924	0.707
2064	18	0.891	0.673
2065	25	0.912	0.764
2067	25	0.954	0.812
2068	20	0.913	0.685
2069	40	0.974	0.686
2070	28	0.991	0.896
2072	18	0.956	0.844
2073	26	0.928	0.741
2074	17	0.902	0.678
2075	18	0.965	0.850
2076	27	0.975	0.937
2077	32	0.988	0.863
2078	29	0.922	0.662
2080	20	0.986	0.918
2081	13	0.969	0.953



426

Table 7

SEQ ID NO:	Chromosomal location
1	2
2	5
3	3
4	5
5	15
6	4
7	12
8	4
9	15
10	13
11	6
12	10
13	3
14	6
15	10
16	12q
17	1
19	2
20	X
21	4
23	12
24	11
25	1
26	16
27	8
28	1
29	11
30	3
31	2
32	1
33	17
34	4
36	X
38	2
39	16
41	19
42	4
43	8
44	4
45	19
46	18
47	6
48	9
49	10
52	11
53	18
54	17
55	17
56	5
57	21
59	4
60	10
61	18
63	4
64	11
65	20q11.21-11.23.

427

Table 7

SEQ ID NO:	Chromosomal location
66	15
68	11
70	14
71	9
72	11
75	1
77	2
78	3
79	7
80	3
81	1
82	13
83	6p11.2-12.3
84	1
85	4
86	5
87	12
88	6
90	2
92	6
95	15
96	10
97	4
98	14q31
99	1
100	5
101	2
102	4
103	4
104	19
105	11
107	3
109	10
111	X
114	X
115	2
116	1
117	5
118	9
120	2
121	19
123	2
124	10
125	5
126	X
128	1
130	3
131	17
135	9
136	16
137	17
138	2
139	2
140	6q16.1-16.3.
142	9
143	20

Table 7

SEQ ID NO:	Chromosomal location
145	8
146	22q13.
147	1
148	6
149	16
151	6
152	6
153	2
155	4
156	17
157	17
158	11
159	11
160	16
161	1
162	17
163	1
164	5
165	15
166	3
168	9
169	6
170	16
171	1
172	4
174	10
175	8
176	6
177	15
178	6
179	9
180	9.
181	2
182	6
183	2
185	11
186	11
188	18
189	11
190	9
191	10
192	4
193	Xq13.2-21.1
194	10
196	20
197	10
198	6
199	11
201	11
203	X
206	8
207	11
208	19
209	15
210	3q
211	6q25.1-26

Table 7

SEQ ID NO:	Chromosomal location
212	9
214	19
215	20
217	1
218	22q13.31-13.33
219	1
220	2
221	3
222	9
223	15
225	3p
226	18
228	4
229	17
230	17
231	1
232	19
234	11
235	19
238	3
239	6
241	11
242	10
243	15
244	4
245	21
246	19
248	6p12.3-21.2
249	3
250	1
251	20
252	16q24.3
253	19
254	14
255	9
257	2
258	11
259	17
260	19
261	8
262	3
263	8
264	16
265	9q34.2-34.3
266	10
267	17
268	4
269	3p
270	9q13-21.33
271	1
272	8
273	19
275	17
279	3q
280	15
281	6

430

Table 7

SEQ ID NO:	Chromosomal location
282	17
283	17
285	15
286	5
289	10
290	9
292	7
293	8
294	18
296	4
297	15
298	15
299	10
300	7
301	5
302	13
304	1
305	Xq25-26.2
306	18
307	2
308	17
309	1
310	12
311	20
313	18
314	11
315	14
316	6
317	10
318	10
319	19
320	9
321	6
322	10
323	3
324	10
325	1
326	16
327	6
328	X
330	4
331	2
332	14
333	2
334	2
336	21q22.3
337	9
338	19
339	15
340	4
341	9
342	10
343	19
344	5
346	16
349	3

Table 7

SEQ ID NO:	Chromosomal location
350	11
352	17
353	18
354	20
356	3
357	5
358	11
359	9
364	2
365	4
366	7
367	5
369	8
370	4
371	6q15-16.1
372	19
374	2
375	12
376	17
377	1
379	19
380	9
381	6.
382	9
383	18
384	18
385	3
387	1
388	21
389	17
390	17
391	4
393	10
394	11
395	11
396	10
397	16
398	13
400	3
402	2
403	Xq28
406	1
407	19
408	8
409	4
410	3
411	4
412	5
413	22q12.3-13.1
414	8
416	8
417	20p12.2-13
418	10
420	4
421	8
423	11

432

Table 7

SEQ ID NO:	Chromosomal location
424	17
425	17
426	17
427	17
428	4
429	2
430	3
431	19
432	18
433	12
434	17
435	6
436	2
438	1
439	8
441	1
442	2
443	11
444	2
446	11
447	19
448	11
449	19
450	3
452	3
453	5
455	17
457	6
459	18
460	18
461	14
462	5
463	11
464	3
465	2
466	11
467	13
470	19
471	6p24.1-25.3
473	4
474	15
475	13
478	8
479	10
480	15
481	9
482	1q23.1-24.1
483	8
484	17
486	15
487	22q11
488	3q
489	1
490	3
492	11
493	1p36.2-36.3

433

Table 7

SEQ ID NO:	Chromosomal location
495	10
496	19
497	18
498	22q13
499	5
501	6
503	1
504	10
505	20
506	3
507	18
508	8
509	1
510	2
513	6q25.2-26
514	6
517	3
518	5
519	12
520	13
521	12
522	15
523	15
524	8
525	15
526	15
528	4
530	8
531	11
532	4
533	17
534	3
535	18
536	18
537	15
538	13
539	8
540	X
542	2
543	5
544	Xq25.
546	11
547	22q13.2-13.33.
549	13q12-13
550	1
552	6q23
553	19
554	1
555	17
556	7
558	11
559	8
560	12
561	10
563	19
564	10



Table 7

SEQ ID NO:	Chromosomal location
565	17
566	9
567	1
568	Xq22.2-24
569	3
570	1
571	5
573	6q22.1-22.33
574	15
575	17
576	5
577	5
578	11
581	22q12
582	16
584	6q25.3-26
585	3
586	11
587	2
588	2
589	15
590	11
591	11
593	Xp11.3-21.1
594	22
595	9
596	11
597	10
598	11
599	12
601	9
602	16
603	12
604	8
605	6
606	11
607	10
608	1
609	3
610	5
611	3
612	6
613	10
614	17
615	11
616	6
617	16
618	11
620	18
621	17
622	17
624	22
625	3
626	19
627	11
629	3

435

Table 7

SEQ ID NO:	Chromosomal location
630	3
631	17
632	6
634	2
635	10
636	12
637	6
639	8
640	5
641	11
642	4
643	7
644	20p12.1-13.
646	15
647	2
648	16
649	8
650	4
651	13q12.11-12.2
652	10
654	1
655	Xp
656	3
657	13
659	1
660	18
661	22
662	X
663	15
664	18
665	4
666	4
667	5
671	11
672	18
674	19
675	17
676	17
677	10
678	10
679	4
680	8
681	5
682	4
683	6
684	1
686	11
687	5
689	9
690	4
691	4
692	5
693	1
694	16
695	19
696	12

436

Table 7

SEQ ID NO:	Chromosomal location
697	11
698	11
699	10
702	5
704	16
705	3
707	3
708	10p11.21-12.1
709	11
710	10
711	10
712	10
714	3
715	6q25.3-26
716	8
718	X
719	17
721	6
722	16
723	2
724	12
725	16
726	19
727	3
728	16
729	6
730	16
731	7
732	11
733	8
734	9q21.11-21.2
735	17
736	5
737	1
738	1
739	1
740	Xq22.3-24
741	17
743	7
744	15
746	12
747	1
748	19
749	5
750	9
751	5
752	9
753	19
754	15
755	8
756	X
757	3
758	1p12-13.3
760	6
761	19
762	8

437

Table 7

SEQ ID NO:	Chromosomal location
763	12
764	2
765	11
766	11
767	15
768	17
769	11
771	11
772	17
773	5
774	18
775	1
777	8
778	16
781	16
782	1
783	21
784	6p21.2-22.1
785	5
787	16
788	7
789	15
790	22
791	6
792	1
793	22
794	8
795	2
796	1
799	6
800	9
802	9
803	17
804	10
805	3
806	2
807	14
810	6
811	10
812	16
813	1
815	16
817	3
818	15
819	Xq22.3-24.
821	1
822	6q16.1-21.
823	17
825	10
826	15
827	3
828	17
829	22q13.33.
830	11
832	15
833	9q31.3-33.2

438

Table 7

SEQ ID NO:	Chromosomal location
834	15
835	X
836	11
837	19
838	10
839	2
840	1
841	8
842	4
843	1
845	16
848	19
849	10
851	2
853	10
856	2
857	1
858	5
859	2
860	19
861	3
862	2
863	11
864	3
865	3
866	21
867	1q42.11-42.3
868	1
870	8
871	6
872	1
873	12
874	6q27
876	11
877	2
878	19
880	3
881	1
885	8
886	9
887	5
888	9
891	16
892	10
893	21
894	5
895	5
896	4
897	13
898	18
899	10
900	16
901	3
902	11
903	1
904	13

Table 7

SEQ ID NO:	Chromosomal location
905	19
907	10
908	5
909	1
911	1
912	5
913	16
914	1
915	8
916	11
917	17
918	16
919	19
920	7
922	9
924	10
925	11
926	11
928	1
929	1
930	12q
931	18
932	15
933	15
934	15
935	1p35.2-36.13.
937	11
938	1
939	15
940	X
942	11
943	1
944	9
946	5
947	4
949	12
951	4
952	10
953	11
956	6
957	19
959	16
960	6
962	16q24.3
963	9
964	6
965	Xq12
966	11
967	15
969	17
970	10
972	10
973	Xq12
974	1p36.11-36.33
976	2
977	20

440

Table 7

SEQ ID NO:	Chromosomal location
979	2
980	8
981	19
984	6
985	5
987	18
988	3
989	11
990	3
991	2
992	17
993	10
994	12
995	1p34.1-36.11
996	14
997	20p12.2-13
998	2
1000	12
1001	1
1002	X
1005	17
1006	1p31.2-32.1
1007	15
1008	15
1009	2
1010	13
1011	6
1012	18
1013	1
1015	6
1016	5
1017	12
1018	5
1019	CITB-H1 2291F22
1020	4
1021	18
1022	1
1023	11
1024	1
1025	3
1027	19
1028	2
1030	3
1031	4
1032	1
1033	3p
1034	X
1035	1
1036	1
1038	13
1041	3

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2535	C	328	546	MMRRPVHCATDKEGILAPKHFQAAAGEA RTSTDRSGAQARSVTPCQWHSVQDSSTY SSVVVVVVAAAAETL
2536	A	163	699	PADAPSLAAFPQDPPYCYPGTQCWV PGEGMLLSQTLCLGEQVLLGAWLVWGFS RDPRPLPYLCHDEPYTFDINLSVNLKGPN RLGEPIPSKAHEHIFGMVLMNDWSGNYW SSVPVKMTGKELGTWGNFIKAEDWCRSK GAVMALPRAVTPTRAINESTIGAAGVDNE VSSTG
2537	A	1415	3050	NHKSPMALPYHIFLFTVLLPSFTLTAPPPCR CMTSSSPYQEFWRMQRPGNIDAPSYRSL KGTPTFTAHTHMPRNCYHSATLCMHANT HYWTGKMINPSCPGGLGVTVCRITYFTQTG MSDGGGVQDQAREKHVKEAISQLTRGHST PSPYKGLVLSKLHETLRTHRLVSLFNTL TGLHEVSAQNPTNCWICLPLNFRPYVSIPV PEQWNNFSTEINTTSVLVGPLVSNLEITHTS NLTCVKFSNTTYTINSQCIRWVTPPTQIVC LPSGIFVCGTSAYRCLNGSSESMCFLSFLV PPMTIYTEQDLYSYVIS*SPRNKRVLPFVI GAGVLGGLGTGIGGITTSTQFYHKLSQELN GDMEQVADSLSLTLQDQLNSLAADVLLQN RRALDLLTAERGGTCLLLGEECCYYVNQS GIVTEKVKEIRDRIQRRAEELRNTGPWGGL SQWMPWILPFLGPLAAIILLLFGPCIFNLL VNFVSSRIEAVKLQMEPKMQSKTKIYRRPL DRPASPRSDVNDIKGTPPEISAAQPLLRPN SAGSS
2538	B	67	1280	XYCRVPTYFHMTPEYEGTTST
2539	A	393	1	GGIGRGGGAGGGVGAAGSASGGVGRRGA GGVIADSGAPGGGVEGGVGASGGWRE/GR GTSGGVGGSGGACGSV/GGSGGAGGGVG ACGSTSDGVGRSRTIGGLGGSGSAGGGV GACGGASGYVGIRGAGGG
2540	A	2	370	ARDPLLEQVELPAVASVSASVIKSPSPSH VSVPPPLLLPAATTRSNSTSMHSSIPSIENK PPQAIKPKQLTHVIEGFVIEGLEPPVRSRS SLLIEQPVKKRPLLDNQVINSVCVQPEL
2541	A	50	247	MWSAHLAVLSLKLTLFSLTSDWLSSKDM AISLAFKISQILCSVLSAPGKRLISVLWNTSS LKRS*
2542	A	130	3995	HPLDIHTILLAAGFLGLRTVGVTKAWRS WLRFPAAAMFLYNLTQRATGISFAIHGNFS GTKQQEIVSRGKILAE LLRPDPNTGKVHTL LTVEVFGVIRSLMAFRLTGGTKDYIVVGS SGRIVILEYQPSKNMFEKIHQETFGKSGGR SIVPGQFLAVDPKGRAVMISAIEKQKLVI LNRDAAARLTISPLEAHKANTLVYHVVG VDVGFENPMFACLEMDYBEADNDPTGEA AANTQQTTLTFYELDLGLNHVVRKYSEPLE



Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				EHGNFLITVPGGSDGPSGLICSENYITYKN FGDQPDIRCPPIRRNDLDDPERGMIFVCSA THKTKSMFFFLAQTEQGDIFKITLEDDEM VTEIRLKYFDTPVAAAMCVLKTGFLFVA SEFGNHLYLQIAHLGDDDEEPEFSAMPLE EGDTFFFQPRPLKNLVLVDELDSLSPILFCQ IADLANEDTPQLYVACGRGPRSSLRVLRH GLEVSEMAVSELPGNPNNAVWTVRRHIEDE FDAYITVSFVNATLVLSIGETVEEVTDSGFL GTTPTLSCSLGDDALVQVYPDGIRHIRAD KRVNEWKTPGKKTIVKCAVNQRQVVIALT GGELVYFEMDPSGQLNEYTERKEMSADV VCMSLANVPPGEQSRFLAVGLVDNTVRRI SLDPSDCLQPLSMQA/LPAQPESLCIVEMG /GT*KQDELGERGSIGFLYLNIGLQNGVLLR TVLDPVTGDLSDTRTRYLGSRPVKLFRVR MQGQEAFLAMSSRSWLSYSYQSRFHLTP LSYETLEFASGFASEQCPEGIVAISTNTLRIL ALEKLGA VFNQVAFPLQ\YTPRK\FVIHPES NNLIIETDHNAYTEATK\A\QRKQQMAEE MVEAA WEDERDLAAEMAAAFNLNENLPE SIFGAPKAGNGQLASVTRVMNPIQGEHTW TLSSLEQNRAAF\SVAVCRFSNTGDDWYV LVGVPKDLILNPRSVAGGFVYTYKLVNNG EKLEFLHKTPVEEVPAAALPFQGRVLIGVG KLLR\VY\DLGKEGSYFRKC*ELRHIANYS GDPDYSGHRVTSDVQEKFHGPFYKRLK KTKLIIFADDTYPRWVHYRPASWDYDTV GWGQDKFRPTYVWVRLPTLTPIDEVR/DE DPTGNKSPVGTGLAQMGGLPRKAEVTEL THVG\ETVLSLQKTTLPGRQLQNSLVLLPP CFGGIG\LVFPVTSHE\DH\DFQ\H\VE\MLR \SEHPP\LCGGGDHL\SF\RS\Y\YFPCEGM*LM GDLCE\QFNSM\EPNKQKERLLKELGPEPPP RSVPRKFEGYSGTRYGF
2543	A	68	425	SHILPGAPGAPAWWTRWPSTLPEPFRGRG SPAGTSPISRPGLVQSS*ASRGSDSRLPV/GP ASCQASGPGPDSRRPPCTPA\GPHHGSLPS AGRVGASAAAAGPPSPA\VLPPAERPAP
2544	A	1	1982	DAERQEALGIVRRIGTDTEAATEPAGATVP AAAAAARIGTVGPQPPAMPRRKRNAGSSS DGTEDSDFSTDLEHTDSES DGTSRRSARV TRSSARLSQSSQDSSPVRNLQSFgteepAY STRRVTRSQQQPTPVTPKKYPLRQTRSSGS ETEQVVDVFSRET\KNTADHDESPPTPTGN APSSSDIDISSPNVSHDESLAKDMSLKDSC SDLSHRPKRRRFHESYNFMKCPTPGCNS LGHLTGKHERHFSISGCPLYHNLSVADECK VRAQ\TRDKQIEERMLS\HRQDDNNRHAT RHQAPTERQLRYKEKVAELKKKRNSGLSK EQKEK\YMEHRQTYGNTREPLENLTSEYD

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LDLFRRAQARASEDLEKLRLQGQITEGSN MIKTIAFGRYELDTWYHSPYPEEYARLGRL YMCEFCCLKYMKSQTLRRHMAKCVWKHP PGDEIYRKGSISVFEVDGKKNKIYCNLCL LAKLFLDHKTLYYDVEPFLFYVMTEADNT GCHLIGYFSKEKNSFLNYNVSCILTMPQYM RQGYGKMLIDFSYLLSKVEEKVGSPERPLS DLGLISYRSYWKEVLLRYLHNFQGKEISIK EISQETAVNPVDIVSTLQALQMLKYWKGK HLVLKRQDLDEWIAKEAKRSNSNKTMDP SCLKWTPPKGT
2545	A	95	719	VWPEVTDPEKFVYEDVAIAAYLLILWEEE RAERGLTARQSFVDLGCNGLLVHILSSEG HPGRGIDVRRRKIWDMYGPQTQLEEDAITP NDKTLFPDVDWLIGNHSELTWPWPVIAAR SSYNCRFFVLPCCFDFIGRYSRRQSKKTQ YREYLDFIKEVGFTCGFHVDEDCLRIPSTR VCLVGKSRTYPYSIEASVDEKRTQYIKS
2546	B	224	429	XPFLILLSPVSTDQANTTTAEIHSQLTPLR NLTLSSQGASLQQRVTYHRNHKYGQTHP QKAEIVVG
2547	A	59	335	GLAAGLPETLHISYCMTVFRFESLDSGVWT DDHSEACRNMHVLSVWTASCKAEPNPIWP HHPWLSCATWPCWKGFDLPGICFTALSCP KIYA
2548	A	1	1605	PMYFLCPPLALVQCALKDPRSKYSLGGR TTLITLQSGGKKNPHPSLSERVMTAKD GFVSRCHLLMQPKQKWSLMYPMEGEVL ENGWCWPTLQDSLLCTALVDKLLVFLGRCF CTAVEVVMVTCRTAAAVSAFLIVGRVSS PVCRAVSVPWTLTADHTPGRYCLKLVCR QLCLCPSSTPLTEVFCSKEAFFIILDCSNLPH ALLPVDSPKGLSKCSNPREKARRKLQGHY HVASEVSFVPVRRFPKGEIGANQPGTHRKF YHLTHYRQNLKQPDVPHGRIVFDDKDITD WQTAKIMREAVAIVPEGRRVFSRMTVEEN LAMGGFFAERDQFQERIKWVYELFPRLHE RRIQRAGTMSGGEQQMLAIGRALMSNPRL LLLDEPSLGLAPIIIQIFDTIEQLREQGMTIF LVEQANQALKLADRGYVLENGHVVLSD TGDALLANEAVRRGDELTEDRSRSLDGELI RSLPCGASYGGLSLRPWSRGHIPQSHQSSE SVRVFMINTSKGASIISSATMPGPLPKHLG P
2549	B	1	597	MHVQGKAAILGRHFSISSLLPGALLLTVIK GHTHPEEKSPGAHEKAVTGEPKCLGALPY CDSGGKKATKKKDAGEMRSRIKDGVLVL KCISLQVGLASWIVSWLRTEATGYTFALLP PGTHHTEQTPSKHEQNGAELFCNCVSCFED PCPCQVPGTQPGNRLSEHQASSQADVTNS SAPKQPHPPPAPCKGVCSHC

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2550	A	278	451	MAGTAQLLGLKQLIGLELLTAQCGQITGY RDRREELLPPRFLATGPPSCHPPSQTPV*
2551	A	1	6530	MWGSDRLAGAGGGGAAVTVFTNARDCF LHLPRRLVAQLHLLQNQAIEVVWSHQPAF LSWVEGRHFSQDQGENVAEINRQVGQKLGL SNGGQELHAVSLEQHLLDQIRIVFPKAIFPV WVDQQTYIFIQIVALIPAASYGRLETDTKLL IQPKTRRAKENTFSKADA EYKKLHSYGRD QKGMMEKELQTKQLQSNVTGITESNENESEI PVDSSSVASLWTMIGSIFSQSEKKQETSW GLTEINAFKNMQSKVVPLDNIFRVCKSQPP SIYNASATSVFHKHCAIHVFPWDQEYFDVE PSFTVTYGKLVKLLSPKQQQSKTKQNVLS EKEKQMSEPLDQKKIRSDHNEEDEKACVL QVVWNGLEELNNAIKYTKNVEVLHLGKV WPKDISEEDIKTVFYSWLQQSTTTMLPLVI SEEEFIKLETKDGPSSRSYGKRRKQGVNSLG VSSLEHITHSLGRPLSRQLMSLVAGLRNG ALLTGKGSGKSTLAKAICKEAFDKLDA HVERVDCALRGKRLNIQKTLEVAFSEA VWMQPSVLLDDLDLIAGLPAVPEHEHSP DAVQSQRLAHALNDMIKEFISMGSVALIA TSQSQQSLHPLLVAQGVHIFQCVQHIQPP NQEQRCEILCNVIKNKLDKCDINKFTDLQ HVAKETGGFVARDFTVLVDRAIHSRLSRQ SISTREKLVLTTLDFQKALRGFLPASLRVN LHKPRDLGWDKIGGLHEVRQILMDTIQLP AKYPELFANLPIRQRTGILLYGPPGTGKTL AGVIARESRMNFISVKGPELLSKYIGASEQ AVRDIFIRAQAAKPCILFFDEFESIAPRRGH DNTGVTDRVVNQLLTQLDGVEGLQGVVY LAATSRPDLIDPALLRPGRLDKCVYCPPPD QDGSSSSDSDLSSMVFLNHSSGSDDSAG DGECGLDQSLVSELMSEILPDESKFNMRYL YFGSSYESELNGTSSDLEDESMNQPGPIK TRLAISQSHLMTALGHTRPSISEDDWKNFA ELYESFQNPKRKNQSGTMFRPGQKFFDEI TELTYPSPFHHKAAPHQAEPGNSSASAP PPYNPFITSSPHTQSGLQFRSVTSPPPSAQQF PLKEVAGAKGIVKTALETAPTLALPVSSQP FSLHTAEVQGCAVGILTQGPGPCPVAFLSK QLDLTVLGSPSCLHAVASAA LILLEALKIT NYAQLTYSSHNFQNLFSFSLTHILSAPRL LQLYSLFVESPTTILPGPDFNLASHIILDTP DPDDCMSLIYLTFTPPHISFFSVPHVDHIW FTDGSSTRPDRHSPAKAGYAISSSTSIIEAT ALPPSTTSQAEIALTRAFTLAKGLHVNTY TDSKYAFHILHHHAVITWAERGFLTQGSII NASLIKTLKAAALPKEAGVTHCKGHQKA SDPITLGNAYADKGVRCAPDPARRPLPLPI GLKACHCSCTAKIGGKYRALVGQLKTISV

445

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				ATGLKTQDRTIDGSSQVIEKNHNGYSVID TGTLVEAELEKLPNNWSPQTCELFALSQAL KYLQNQKTISILIQKEPSALGLTPERKGNV GHAGKGPLESSSPDPFLCGQERREKGCRTA TSVSITNPINRGPWVVTHPGKELTPEHKGN VGHAGRDILAKAGAIHLNIGEGTPVCCPL LEEGINPEVWATEGQYGRAKNARPVQVKL KDSTSFYQRQYPLRPKAQQGLQKIVKDL KAQGLVKPCSNPCSTPILGVQKPNRQWRT LCHQATQALFNFLATCGYMVSKPKAQLCS QQ/RYLGLKLSKGTRALSEHIQPILAYPHP KTLKQLRGFLGVIGFCRKWIPRYGEIARSL NTLIKETQKANTHLVRWTTEVEVAFQALT QAPVLSLPTGQDFSSYVTEKTGIALGVLTQI RGMSLQPVAYLTKEIDVVAKGWPHCLRV VAAVVVLVSEAVKLIQGRDLTVWTSHDV NGILTAKGDLWLSDNHLLKYQALLLEGPV LRLCTCATLNPATFLPDNKEKIEHNCQQVI VQTYAAQGDPLEVPLTDPDLTLCTDGSSFV EKGLRKVG YAVVSDNGILESNPLTPG TSAQ LAELIALTWALELGEEKRANIYTD SKYAYL VLHAHAATWKEREFLTSERTPIKHQEAIRK LLAVQKPKEVAVLHCRGHQKGKEREIEE NCQADIEAKRAARQDPPEMLIKQPLV
2552	A	748	1075	ILPTSLFFLCFVFFVCF*DRVLLSPG\WSA VARSWLYCNLSLRGFGFSCLSLSNWDY RCTPLRSANFVFL/CRDRVSPCWPTSVSNS* PQ\VIHPPWPPKVLGITRV
2553	B	1	766	MRPVDPDGTEHSLFCPLTALRGMVNSRIQ KSPGKPSVCDVPLPISPGQSSQLHGKVFQ LNAGKAAEFKSPDPHQAAASTSGPQKT TLSKRGLRLQPCQLHSAPHSFQLPLTQKS TWDLRGSAPLHAAQTSLSFESCHRPDVED TLGTKGPDKTQCQSENSTRPQYSPETSQNQ PVGKGTDLKVTKLGVPSLMAQDGVNYSV KTEAHSTGTTAEPLSSQDRAVRGHNTDSH VQTPDLGEDTAL
2554	A	47	923	KATRFISAAFVVLNKQGVSPAKLPHTSWS WSLQTLFSLFSGDLAEKSLQCFPCSAMLE LIPLLGIFVLR TARAQSVTQPDHITVSEG ASLELR CNYSYGATPYLFWMERTVEEAFIL LVCLKPWRVASSLEKKEKEDES FQLLGSR YNVLKGSRGETSEGGAESFSSQSPGENQLY SEMQFFYLCEQRAVVPTESWVGLNLFM ASWMKHSGKLWSKRNSEELCGTLHITAAQ LKDSGTYFCAVEAQFSQEICSLDPNC SWAC SPNPFRRERGM LPPQYHLHSFGFSD
2555	A	2471	2985	ETSLERERLSFCTGSRTTRSAELKAVGF EA ALQEVITPEVVPASQSEA YQTLRQNAQV HNFFFWGGDSP T LSPRLECSAISAH CNLR LPGSSNSPTSASRVAGTTGACRHARLFCIL

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				VEMGFHRVAQAGRELLSSANPPTSASQSA GITGMSHHAQPSSQLLISSCC
2556	A	138	564	YREVMVSES*ETPAGARGRPYYFSAPGTAP \PAINVHPPPPSLSATPHPPQPQPPPHQHNA KARVATIRTKRTSNCRIRSRKVRKSPPEKW VGFNRPRKASCSPPPGAARVDVGGETERR EQAAAPGEMGKWARPGEYFHS
2557	A	2	585	AAAAPAGGNPEQRLDYERAAALGGPDGR AWGGRSPLPPAP*AQGAPGRWPPPRAGS PAPSPAGCGGGKGGGLVTPGRGGPRAAGR EL/RAVRCPCPVRPRPPSKPALGGSPLQPEP AAAPGPSIR/PVLPITQTS/PWRRPKSLRPVL GTRVGRTPPLPPP/DPAGPPPLPLPGP/HPS RPPPTGPWRPARADGRV
2558	A	2	224	PRVRVQWAQLSQDKKGEMNSMTSTAGPP GSSSAPCATRRNLLQRQHLQRLSGEFKKDP ATYSKHLEPLEEERDK
2559	A	43	267	GRLWSAMTPGKLTCLKIDWPALEVWGP LEGSLDRSLVSKVWHKVITYKPRNPDQFPY RDT*LELVLDPPPPTHSG
2560	A	233	692	DNHPSFRLPSSRPGTKEVLKEIHISDTTAD VIFYPIYRMSEMIFRRIKMPWLWDLWYL MFKEGWEHKKSLKILHTFTNSVIAERANE MNANEDCRGDGRGSAPSKNKRRAFLDLL SVTDDEGNRLSHEDIREEVDTFMFEVLYTV RFRYH
2561	A	1993	1379	SLHLSERADWQYSQRAG/DAVEVFFSRTA RDNRLGCMFVRCAPSSRYTLLFSHGNAVD LGQMCSFYIGLGRINCINIFYDYSYGVS SGKPSEKNLYADIDAAWQALRTRYGVSP NIIYQGSIQTVPTVDLASRYECAAVILHSP LMSGLRVAFPDTRKTYCFDAFPSIDKISKV TSPVLVIHGTEDVIDFSHGLAMYERCPRA VEPLWVEGAGHNDIELYAQYLERLKQFIS HEL PNS*RQSK
2562	A	991	308	AAASAFKPLGLALSDRAFAAWEPGAAVSR SPLSPSRPFASREPAGFRAALADPPGMPR YELALILKAMQRPETAATLKRTIEALMDR GAIVRDLENLGERALPYRISAHSQQHNRGG YFLVDFYAPTAAVESMVEHLSDIDVIRGN IVKHPLTQELKEWEGIVPVPLAEKLYSTKK RKK*EDSPDFSLICNSFTFGQHGREGRICKF GLYISMCCRCLIFLRYF
2563	A	1	344	MDKSLLELPILCCFRALSGSLSMRNDV IEIVQCRMCHLQFPGEKCSRGRGICTATTEE ACMVGRMFKRDPNPWLTFMGCLKNCAD VKGIRWSVYLVNFRCCRSHDLNEDL
2564	A	251	386	LQRLECSGTI/SAHCNLCLLGSSNPLASAS*I AGTTGTGTGDVDST
2565	A	1164	1273	EISNIQQADFPGLATHPAFSRLLPCLHFIP KSAHQ

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2566	A	867	156	PAPVKDEGPMVSASVKDQGPMVSAPVKD QGPIVPAPVKGEGPIVPAPVKDEGPMVSAP IKDQDPMVPEHPKDESAMATAPIKNQGS VSEPVKNQGLVVSGPVKDQDVVVEHAH VHDSAVVAPVKNQGPVVPESVKNQDPILP VLVKDQGPTVLQPPKNQGRIVPEPLKNQV PIVPVPLKDQDPLVPVPAKDQGPVPEPLK TQGPRDPQLPTVSPLPRVMIPTAPHTEYIES SP
2567	A	625	182	QQGKNQECIRNQHTRAPGRGASPPQGEGK TWAUVGHPVPHALVIPGLQSGSARGLA RQLGRAR*PRPPAPPRACRPEEPYTPGRR APGRPAPAPRSACGWAASASRWCRRTVFF SQ
2568	A	2	917	EELCLDVSENRLERLPPEISGLTSLTDLVIS QNLETPDGIGLKLKLSILKVDQNRLTQLP EAVGECESTELVLTENQLLTLP*SIGLKLK LSNLNADRNLKLSLPKEIGGCCSLTVFCVR DNRLTRIPAEVSQATELHVLDVAGNRLH LPLSLTALKLKALWLSDNQSQPLLTFTQDT DYTTGEKILTCVLLPQLPSEPTCQENLPRCG ALENLVNDVSDEAWNERAVNRVSAIRFVE DEKDEEDNETRTLRLRATPHPGELKHKMK TVENLRNDMNAAKGLDSNKNVNHAI VTTSV
2569	A	481	1380	TSKQNAAPLVKYFQEKGLIMTFDADRDED EVFYDISMAVDNKLFPNKEAAAGSSDLDP SMILDTGEIITDGSYEDQGDDQLNVFGED TMGGFMEDLRKCKIIFIGGPGSGKGTQCE KLVEKYGFTHLSTGELLREELAS*SERSKLI KDIMERGDLVPSGIVLELLKEAMVGSGLD TRGFLID\GYPREVKQGEEFGRIWRPHS WVICME\CSADT\MTNRL\LQRSRSLPVDD TTK\TMAKRLEAYYR\ASIPVIAYYETKTQL HKINAEGTPEDVFLQLCTS*LTLFSEGKN ACLG
2570	A	3344	677	GAYHKHLMELALQQTYQDTCNCIKSRIKL EFEKRQQRLLLLSLPAHIAMEMKAETIQR LQGPKAGQMENTNNFHNLYVKRHTNVSIL YADIVGFTRLASDCSPGELVHMLNELFGKF DQIAKENECMRIKILGDCYYCVSGLPISLPN HAKNCVKMGLDMCEAIKKVRDATGVDIN MRVGVHSGNVLCGVIGLQKWQYDVVSH DVTLANHMEAGGVPGRVHISSTLEHLNG AYKVEEGDGDIRDYPYLKQHLVKTYFVINP KGERRSPQHLFRPRHTLDGAKMRASVRMT RYLESWGAAKPFALHHRDSMTTENGKIS TTDVPMGQHNFQNRITLRTKSQKKRFEEL NERMIQAIDGINAQKQWLKSEDIQRISLLF YNKVLEKEYRATAPAFKYYVTACCLIFFC IFIVQILVLPKTSVLGISFGAAFLLLAFILFVC

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				FAGQLLQCSKKASPLLMWLLKSSGIIANRP WPRISLTIITTAIILMMAVFNMFFLSDSEETI PPTANTTNTSFSASNNQVAILRAQILFFLPY FTYSCILGLISCSVFLRVNYELKMLIMMVA LVGYNTILLHHAHVLGDYSQVLFERPGI WKDLKTMGSVLSIFFITLLVLGRQNEYCYC RLDFLWKNKFKKEREEIETMENLNRVLE NVLPADVVAEHFLARSLKNEELYHQSYDC VCVMFASIPDFKEFYTESDVNKEGLECLRL VLEHADFDDLLSKPKFSGVEKIKTIGSTY MAATGLSAVPSQEHSEPERQYMHIGTMV VEFAFALVGKLDANKHSFNDFKLRVGINH GPVIAGVIGAQQPYDIWGNTVNVASRMD STGVLDKIQVTEETSLVLQTLGYTCTCRGII NVKGGKDLKTYFVNTEMSRSLSQSNVAS
2571	A	3222	5798	PLLTPLVSKVTAAGVPLFFFFF*DIVSLC HPGWSAVV*P*LTAASNS*VKQSSHLSLPS SWDNRYAPPRPANYFYFYFL*RLDLALFP KLLNCWAQVILPSQPPKVLGL*AQSSEGG IHGSLSLPSPCFLLCNPI
2572	A	1	666	ASSTPQVTANEEINVTSTDSEVEIVTVGESY RSRSTLGHSRSHWSQGSSSHASRPQEPNRN SRISTVIQPLRQNAAEVVDLTVDDEPTVV PTTSARMESQATSASINNSNPSTSEQASDT ASAVTSSQPSTVSETSATLTNSNTTGTSGD DSRRTTSSAVTETGPPAMPRLPSCCPQHSP CGGSSQNHHALGHPHTSCFQQHGHFQHH HHHHHTPHPCI
2573	A	300	110	PCGPPQEKGADCHLKACPTAPCTTFRASCC SHPASCSRGRKQASMSSTSSATVPLPANEM HSG
2574	A	2	362	QELERSMAQRCVCVLALVAMLLLVFPTVS RSMGPRSGEHQASRIPSQFSKEERVAMKE ALKVFPTVVSTSFQHEVVEEYSHLFTIQGS DPSLQPYLLMAHFDVVPAPEEGWVPPFS G
2575	A	1740	2026	ENGSLRPKPTGIPLSSARGNELSPTRRRRRP WTPNPAGETMSSVQQPPPPRRVTNVGSL LLTPQENESLFTFLGKKCVGAGRGGRAPPS RAAGE
2576	C	363	692	MLLWPLTQAQSSEMSCRLGACFITSLLHQ IPATALLEGNDITLTVQLQILDHNPFPYRL CLIDRCICFISSTYPQIDGLKSSRDIGDKISF VRNGSINMGKPFNF
2577	A	1	2169	MEGLNWLSLLAFIFLLCWMLSALKHQTPN SSAFGLLDLHQWFATGSRMNKNNKPSSFI AIRNAAFSEVGIGISANAMLLLFHILTCLLK HRTKPADLIVCHVALIHILLPTEFIATDIF GSQDSEDDIKHKSVTYRRNRQSQHFSHTNL SPKAPPEKMATQTILLVSCFVIVYVLDV VASCSGLVWNSDPVRHRVQMLVDNGYAT

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				ISPSVLPRLTAPNEWRASVYLNDNLNKCNSN GRLLCVDRLDEGPRSVPKCSESETDEDYI VLRAPLREDEPKDGGSVGNAAVSPEASA EEEEEREEGGEACGLERTGAGGEQVDLGE LPDHEEKSNOQVAAATLEDRTQDEPAEES CQIVLFQNNCMDNFVTSLTGSPYEFFPTKS TSFCRESCSPFSES VKSLESEQAPKLGCAE EDPVVGALCGQHGPLQDGV AEGPTAPDV VVLPKEEEEKEEVTVDDMLANPYVMGDEGE EEEEEFVDDTLANPYVMGVGLPGRGEEEE EEEEVVDDTLASLYKMGEHRHKGGLAPL WEGGQKPSQKLPPKPDRLRQVPQPLASEV PQRRQERAVVTEGRPLEASRALPAKPRAFT LYPRSFSVEGQEIPVSISVYWEPEGSGLDDH RIKRKEEHL SVVSGFSQRNHLPSSTSTPS SMVDIPPPFDLACITKKPITKSSPSLLSDS PDKYKKKKSSFKRFLALMFNKMERP GTM AHACHPSTLGS
2578	B	1	360	MHLLQAALLLAVPCLLCYVAVGYAFSVLL TLLLTAPALLPDDFEGFNIREKTGWYKKE GMVTLNPNQVAREKEQFNDLYFNAKQAE QKGYLNTARREASLAFKVTETTHNKSGLIT ES
2579	A	1	1036	ATVGGREIYVKGFVHYKVRALFPCEKPPRP TEMSRHHSRFRDYRVGWDRREWSVNGT HGTTICSVTSGAG/ERHSQQQPARPPAA ARGALPAAHPGYSSCSL/RPPAAARPSPAS WPALRLRSPRLPASPKGTVSPRDWRPASG GGRRLSISPHPG/ITDEPPSKQMRESNPGT GPW/GPRWPPGTSP*SHTPMEWPSLPPSP GCERP/GPHWGDPLTASPRGAPAPADARP L/PLPQPPSQPLSS/GWSTCLPRCPMPALSP WPCPHCPVWGRWPAQDPPLWATATWQG PCCLHRRQPSRPLSPVPLPPMGPPQPTRP TGCRCGPLAWGSMSSPTRGTPE
2580	A	1	1535	MEEKTNVQLPPGQTEQHVEIHIMNFCSKN HHRITPEKPKELTDPFKEAACCKLYEIDK KLYRMAEWIKIHKPSICCLQETHLTHKDSH KLKVSITFKDLAVRFSEEWRLLEEGQREF YRDVMRENYETLVSVEPGRAVGGGSHAD EGQEPAGCG/VSPGPAAAGEGDPRLVWR SQGRYGQPRER/GRGASLDGERASPEAA/D GKRALPSRPAQLPSRRPYQPAPPGPTPTD SSCSSGPTGDGVQGSPLPIRISPGNSPL/PRP HQLSEGNPCAWAPAPRDIPKLLATSP*PGH VQANQSRPGAWEPALGRSDQRACSASGSA ELCERWPQQAP/APPEEPPASPHPAAPTG/ PGFWESCGEPAAPGKGSAPKPSPLHCLE SALRGILP/EGPCASPAWEAPAPAPAPAPAR ASAA/AEGEDPRPEPELWKPLPQERDRLPS CKPPVPLSPCPGGTPAGSSGGSPGEAPGEO



450

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				SPGTAAASVQ/VSPAHWPCFS/SPVRYSSGS LPGFSAGEKAQG
2581	A	3	514	PRLMEAGPHPRPGHCCKPGGRLDNMHGF VHHIRRNQIARDDYDKKVKQAAKEKVR RHTPAPTRPRKPDQVYLP RHRD VSAHPR NPDYEESGESSSSGGSELEPSGHQLFCLEYE ADSGEVTSVIVYQGDDPGKVSEK VSAHTP LDPPMREALKLRIQEELAKRQSQH
2582	A	307	1503	GGSSARPRASSRRLSRKKT KNEVSKPAE VQGYVKKETSPLLRNLMPSFIRHGPTIPR RTDICLPDSSPNAFSTSGDGVVSRNQSFRLT PIQRTPHIEMRRESNRLSAPSYLARSLADVP REYGSSQS FVTEVSFAVENGDSGRYYYSD NFFDGGQRKRPLGDRAHEDYRYYYEYNHDE QRM PQNQGRHASGIGRVAATSLGNLTNHG SEDLPLPPGWSVDWTMRGRKYIDHNTNT THWSHPLEREGLPPGWVERVESSEFGTYIV DHTNKKAQYRHPCAPTCTSV*STTSCHI/A S/RQQTERNQSLLPANPYHTAEIPDWLQV YARAPVKYDHILKWELFQLADLDTYQGM LKLLFMKELEQIVKMYEAYRQALLTELEN RKQRQQWYAQQHGKNF
2583	A	1341	1015	LGTRGCLNMAAPLSVEVEFGGGAELLFDG IKKHRVTLPGQEEPWDIRNLLIWIKKNLLK ERPELFIQGDSVRPGILVLINDADWELLGEL DYQLQDQDSVLFISTLHGG
2584	A	1	741	VRSMSCPPSWPYCAPCPTNIGESTSPLRKT ETPTLWDPKAPSCSLELPWVLASQRSRG TALPFLPSNVLPALPSTSFCLRP LSHLV TSLLAGPGAHDGHLRKEGWRSTPEMTSLP APEHPASPCDSVLCSPDVSMCTLGPAARW DAQAKSAPLPPCCTDCKSFPHLQRPWAQP HTSQATSVDSEAGTKGMSQFTVWTWWR SRPCETRQGEIGNWGYSVTPGPPGSQNL ARLDGQGLAS
2585	A	36	363	NAHSLPIEWAFCKIENLCGKCVYMCMCSQ NKNNQLKFSFIPGRWCASLKMYSKGQRSL MYPCRYHQRM LLSRYLDTVLLDWDPPG PLPEGRQHSPGRRQRDLASALLC
2586	B	1	1107	MLYWLMPKGKLLWIASFLTRLQGIQHTLP RVEEKSIQSVKDDNIYHPRPRIAVVGSSS TVISYSPGEYAFTNGTSRCP SLSLAAGPRLI TNGPWEAHEVQRESTIALMKLLQVLEQKV RLREGHSLGTVKMSKNINPMGHVSNPPTS YPDELITKQVCPGSHPKRPGEVKHNEEVPT SQDRDTCTTQETQYSVRKIISAEDDFTVKN YNHIRNKFTIPSRKGQQAHRWLNKAIPQP MPTSATSLAALVRAAKHRNQPPQDLAQS SSHHIYLFITITFGSLRDELKSKRGPDQLS LELEMVAKAKAVKPENSRRWFSGNQLGSI INSPKKGSAVLEGTQFEKQKWDARLTKGD

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				CLATNVLNRV
2587	A	1	384	MACRVLQGLPFACLSPPICSHSALTDHSL LCLNFFFNPYALLPNFYLFQKFLRIPSSGK PFLTSEDQDPAGIFELVEVVGNQTYGQVY KVRMRVWKYDLHCRAVLGGVEGSRFLV CRSEGGYGRC
2588	C	1	417	MLLPLFLLIHTGIGPSYSASDRAEPRSPGG RLTARIWIKGVKEDGGTMQGAVDWGEGV ERCAGRITASHKVADKWHSSRNSLGGSP PGTPAPGPVWGFCHPCLPASPLSWTATGT AATHAQCAERVHNLCCRRAKPS
2589	B	1	198	MQAGLARAMVLAAGWSRVASAGAAGDT SPVPRALSDLRITQKCGLLVPKAVSWKSLF LFPITVEL
2590	A	267	614	MAVAVLLCGCIVATVSFFWEESLTQHVAG LLFLMTGIFCTISLCTYAASISYDLNRLPKLI YSLPADVEHGYSWSIFCAWCSLGFIVAAG GLCIAYPFISRTKIAQLKSGRDSTV*
2591	A	5	447	SSAFRSVLLEMVRSSRTCHDTLQGAVPITYP GSGTPALGEKSGSLGLVAWSFPRPGESSST APRRSPCCCPWSPSHSSPASFPPLRPSAPAT RAPREGLPTPASRAHFGATAIPKTSGLLIA TASLCWGQTHQPCPLPLARFLGKR
2592	A	508	870	GHCPVLRVVTEKHCRACEKEGMDSSIHLS SLISRHDDEATRSTSTEGLEEVEGETLLI VESEDQASVDLSHDQSGDSLNSDEGDVSW MEEQLSYFCDKCQKWIPASKELLNSFDLSI PV
2593	B	20	201	MGRVSGLVPSRFLTLAHLVVVITLFWSRD SNIQACLPLTFTPEEYDKQDIHALPAVTEM ALFVTVFGKPKPF
2594	A	79	243	MSFICFLNFVPTSAIPLRLWNYCGMNSPS RSWDCLCTPLSRQSAPVSHMAKVV*
2595	A	178	1224	RYRAARNVMKDQRLVFHRSKVRSSGYASA PHVTMFSPKTNKSEKGGSSRSRSCAREA YPVECAVPTKPGPQVAAAPTCTRVCCIQYS GDGQWLACGLANHLLLVFDASLTGTPAVF SGHDGAVNAVCSQDRRWLLSAARDGTL RMWSARGAELALLRYKQKSKSLICRLST TGAVDMTSLAVNDFYSHIVLAAGRNRV EVFDLNAGCSAAVIVEAHSRPVHQICQNK GSSFTTQPPQAYNLFLTATAIGDGMRLWDL RTLRCERHFEGHPTRGYPCGIAFSPCGRFA ACGAEDRHAYVYEMGSSTFSHRLAGHTDT VTGVAFNPSAPQLATATLDGKLQFLAE
2596	A	85	839	RSGSLMAAAAATKILLCLPLLLLSGWSRA GRADPHSLCYDITVIPKFRPGPRWCAVQGG VDEKTLHYDCGNKTVTPVSPLGKKLNVT TAWKAQNPVLRVVDILTEQLRDIQLENY TPKEPLTLQARMSCEQKAEGHSSGSWQFS FDGQIFLLFDSEKRMWTTVHPGARKMKEK

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				WENDKVVAMSFHYFSMGDCIGWLEDFLM GMDSTLEPSAGAPLAMSSGTTQLRATATT LILCCLLILLPCFILPGI
2597	A	319	513	IELRAVAQGIAQSLGQLLFTQCPLKDDLE GLFLQNNKEGVQKGRDEPLPLP*ATALSS IQAGIQQAR*EGDLEAWQFPVRIHPPDQGG NIIVTFEPFFKLFKEFKQAVNQYGPSPFV MGLLKNVAVSSWMIPTDWDALTRACLTP AQFLQFKTWWADEAGRV
2598	A	1257	877	AVFTFHNHGRTANLYSLHSWLGITTVFLFA CQRFLGFVFLLPWASMWLRSLLKPIHVFF GAAILSLSIASVISGINEKLFSLKNTTRPYH SLPSEAVFANSTGMLVVAAGLLVLYLLAS SWKRP
2599	A	54	470	CSTMNPSEMQRAPRRQRHRSRAPSAHK MNRVMVMSEEQMKLPSTKKAEPPTWAQLK KLTQLAKKKLENTKVTPENMLLAALK TVSTVSAGVPSSEESDHRERAMMTTVVL SKRRGKCGEKKEISDCYCVYVERS
2600	B	1	939	MALRLVIPALWEAELVGALMLAALSHLHR FLLSMWVLPPTFTDAFFGLLFHPPRRSQK DCLLGLSKSDQRAMACYFGILLIVSATLCF GMNYYLDEFANLLDELLMKJNGLSDSLQL PILLEKTSNNTGEARTEESPLVDISSYQAAE MVM MARTLATCLQHAQGLGFEACLPILSA PHALSHWTLTTCLWQLGFMSAVLILKYTR ALLAQGFSGPFVIDKGVRELI GLISRVW EVSEQENSKEEVYRHEEGITVISDLLGRQ WQQGHKGICLQLMLPFSRGKHRTSGAFLM FSLELFTVAQLVPISGS
2601	A	1	698	VLNPLGKP*HDTPAWHEEGYPPTAPPVDP FAKIKVDDCGKTKGCFRYGKPGCNAETCD YFLSYRMIGADVEFELSADTDGWVAVGFS SDKKMGGDDVMACVHDDNGRVRIQHFY NVGQWAKEIQRNPARDEEGVFENNRVTCR FKRPVNVPRDETTVDLHLSWYYLFAWGPA IQGSITRHDIDSPPASERVVSIYKYEDIFMPS AA YQTFSSPFCLLLIVALTFFYLLMGT
2602	A	2	319	FYLFILFLFFVFLVETGFHHVQAGFELLTS SDPSALASQSARITGMSHHA WPNFCLLSRD QVSPCWPGWS*TPDLR*STFLGLPKC*LQA *ATVPSAGEPQCGQ
2603	A	147	773	MGLGARGAWAALLLGTQLVLALLGAAHE SAAMAASANIENSGLPHNSSANSTETLQHV PSDHTNETSNSTVKPPTSVASDSSNTTVIT MKPTAASNTTTPGMVSTNMTSTTLKSTPK TTSVSQNTSQISTSTMTVTHNSSVTS AASSV TTTTMHSEAKKGSKFDTG SFVGGIVLTG VLSILYIGCKMYYSRRGIRYRTIDEHDAII*
2604	A	2	331	WVFSSPITARDALGIKHTMVKIRPLSQATR AAKAKARAYAEFLQPAKERPETS AALARR

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LVISALGVRSKQSKTEREAEKKLQEARER KRLEAKQREDIWEGRDQSTV
2605	A	549	641	CCCCCCCLCFGIHSSKGTHSANSKDWPFDP
2606	A	1	517	SCYVCGGTVTGDQWP*EARELVPTDPVPD EFPAQKNHPDNF*VLKVSIIHQYCTAIEGKQ FTHSIGRLSCLRQKLYNGTTKTVTWWNSN YTERNPFSKFPKLQTVWAHPEFHWDMWA PTRLYWICGHRAYAKLPDQWTGSCVISTIK PSFFLLPIKTGELLGFPVYASHEKR
2607	A	2	406	FLVETEFVVGQAGLELLTSRDPASASKG AGMTGVSHQVQPQ**S*LWT*/PSSVEAGT SFGLSFLSSSWALSAQEGCLA VPS/SGSRGL LVGALLWTKPSPQLSPVPASQRLSSLSLM PPLPQPQHLTHTSIET
2608	A	2264	37	FFFNKNLLFIQKLTGPVFSPIFKKKKKRGGQ GFPSQCP*VNSLAIQGWPSRGVSGKRCQKC GGPGPLRTHSPLLASPLQPPS/WTRPVGLQ PPGAL/GLTTTRGRAALP*LP*N*MLKPRW EQGDFPPGGWAMEAFSRDSLPLQEGIPGP TSPPTPSEK*INKVPETPGALV*ETGCQTEKH FRGGDVSTEGDTYACLDVILNVACLDHGK SEHSPKSPSTQSEEQTLRGRGQAVADWPPG AGACPGPSARLCRGTMGMPSASEHLKRAA LGGK/PPLWRGARAAQEA PGSGFCGITAAR GLGRGGGRDRSLPGKL**KWPVSSTPPGPG RAALPAALGWPGCGPTGM/PGLRSASIPSA KARSHTCGFKPG/LKGRTMEEGQTHRRG PHA*AQTPSATGQVVQQC/PVPLDQRGKSS LRQRPKESNLTGKDLPHPLSPKPPCARSLPQ TPGQSPAELQPLVLSRSPGPAAEQGAD WQGPQRIHPSKWPVKVEPLTPSLQDVGGG GGVTVGACSPRGLPMNASGGTLGLAEC SQGEQPRSPTRQRHHGRGLPRAGLLAEG GNRGPKC/PPLKHGLMGC*LCKAAARILD GLALTVWEAASHPSLPCARTPSGSQRALK GLGGTRKCCGKGQGVPHD\NSSAGTDP THQQPRNRGCA/GDS DSPSGCWGQANLT TAS PATGN*TPGLE*HDVGM EKGLQDQ/Q PGPP RSADGATETQRGQEA AHNQRARGRT LGS YLWSRVGSHSW
2609	A	1	399	MDGQARWLTVPVIALWEAEVFIEHMLYAL NILRTVLGRARTLSLNHRCLLLSLLVLH CVRSVRSWYLFCEAAAEKTLAFAMAEK KALSMGQIRFRFDSQPINETDTPVQVEMED IDIDVFHQQIGGVY
2610	A	1	1641	MGBLHMITEEKHQPFMDTQTAAKGTLL EAGPLDPVCLGHIKKVIQRKFWRYSA PGTVP TTSAPGETEWGRLPQWSTAW SETAQHGW PAARQSRITVLHQQPQCD PGPEVTSEQLPG VINMLTLKYIKVAA HPHGSWNTRVPCLV VLLTPRLSY YIIEIQTTFREYKYHLYENKL

454

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				ENLEEMDKFLDTYTLPRLNQEEVESLNRPMTSSEIEAVINSLPTKKSPGPDGFTAIFYQRYEELVPFLLRLFQTIEKEGILPNSFYEASIIIPKPRDRTTKNENFRPISLMNIDAKTLNKIMANRIQQHKKLIHNNQVGFISGMQGWFNICKSNIHHINRTNDKNHMIISIDAEKAFDKIQHPFMLKALNKLIGDTHLKIIRAFDKPTANILNGQKLEAFLKTDTRQGCPSPLLFNVLVLARAIQKEIPAPADTSSIAHHPSPSYQPWTPVTRTSHSTPTTTCYPCLECTPAKWLTSTVMGGGLLSVPQGTVRVSALNYCFIPQLGGGPLMASSASSDYVPESDESEPLTFE
2611	A	146	411	LLSPSHPLTAPPPRPPPPTRAPGACASSMGPPTSKFPKDLTLPDAALGCGTPATGGEGASSRARSETQARAPTPGRSWGRAGSA
2612	A	2	384	PICLFSRPTLRPSRSKVSLEGRGANMAARWRFWCVSVTMVVALLIVCDVPSASAQRKEMVLSEKVSQLMWETNKRPFVIRMNGDKFRRLVKAPPRNYSVIVMFTALQLHRQCVVCKYELQLRFFIK
2613	A	1	626	SRVEDFVLHLLRALAQDDVVPYFKTEFGLPQIHLEGNRLVLTCLAEGSWPLEFKWMRDDELTTYSSEYKYIIPSLQKLDAGFYRCVVRRNMGALLQRKSEVQVAYMGFSFMDTDQRKTVSQGRAAILNLLPITSYPRPQVTWFREGHKIIPSNRIATTLENQLVILATTTSDAGAYYVQAVNEKNGENKTSFFIHLASFCGNTTQD
2614	A	412	1	SNLCLGNSWRWRWAKSRHHCIPTVTLSKRSGDIRGSHFSSPQRQSRQVPGKETARVLRAGKQGRGQIPCPWPPPPPPPPGSPGCRQFHQSLEAKARHPASVREMRGKVKMRRALRRAPASTRASSRQPNPK
2615	A	2	474	TGPTIKNMDGTFNVTSLKLNSSQEDPGTVYQCVVRHASLHTPLRSNFTLTAARHSLSETEKTDNFHSIHWWPISFIGVGLVLLIVLIPWKKICNKSSAYTPLKCILKHWNFSFTQTLKKEHLIFFCTRAWPSYQLQDGEAWPPEGSVNINTYSTTV
2616	A	223	2210	SLSGFTREASFEMAAQRIRAANSNGLPRCKSEGTLIDLSEGFSETSFNDIKVSPSALLVDNPTPFGNAKEVIAIKDYCPINFITLKFSGKDHLVLDTSGGEWYAHNTTEMGYIPSSYVQPLNYRNSTLSDSGMIDNLPDSPDEVAKELELLGGWTDDKKVPGRMYSNNPFWNGVQTNPFLNGNVPVMPSLDELNPKSTVDLLFDAGTSSFTSSSATTNSTGNFDELPTNGLHAEPVRRDNPFRRSKRSYLSLSVLQAKSDAPTSSSFTGLKSPAPEQFQSREDFRTAWLNHRKLARSCHDLDLGQSPGWGQTQAVETNIVCKLDSSGGAVQLPDTSSIHVPEG

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				HVAPGETQQISMKALLDPPELNSDRSCSIS PVLEVKLSNLEVKTSIILEMKVSAEIKNDLF SKSTVGLQCLRSDSKEGPYVSVPLNCSCGD TVQAQLHNLEPCMYVAVVAHGSPSILYPST VWDFINKKVTVGLYGPKHIHPSFKTVVATIF GWHDCAPK\TLLGSGEVTRQAPNPAPVALQ LPQDLKVCMFNMTNYEVKASEQAKVVR GFQLKLGKVSRLIFPITSQNPNELSDFTLRV QVKDDQEAILTQFCVQTPQPPKSAIKPSG QRRFLKKNEVGKILSPFATTKYPTFQDRP VSSLK
2617	B	10	462	MSGWLGLVSSLHRLLVSPCPGRTVGLQRR KRLKSGSSRMSFPVTRRPREQTPHPDIVAAI PSGTDDFQGHRSKEKENWKPMCLNRFILE ECIAADDFRIRGLEPNPQYLQGKPTQVSES LRLLRNDTQDPNIKTRYIMNLAktiQIRSPD K
2618	B	1	406	MIIPKLNLMCALQSKPESRGFGELSQRGN VKFNVTLCSHQKKISRLSAAIHQLDISDIR PLTVLLTLCITLALLMRGAQPGMNSGKIPY RMFIPNSHSDSELMSFQDSVRHRRGGFQTF DCDSQQETFWTWSIX
2619	B	1	789	MGRERDPGSGWTWLLRCAAAACALLGSQ RQETQLLLSEHSDPDIEHRVRGEPKRTTRW LGVECWRQGVINIETKAQEQLQPKGKKVS SLLTALPGSIDELSLKRDVKESISLPAVPFQI ELLISKINMQTRLLQLPLKFAVAAASSRF NPRPPVIGQLRGKKSTPWQDPKPIKSPAG VTAATLQAGVGWABEQSGHCAQVHSLGV DSSCWSRSGYTYVHHPVHTPTLCALVGS GGERGGGEGEKHIGLEEQEPQKRVLN
2620	A	3	913	FMTDVNSWLLTFGFQLHNVIPGYPKPDM AMEPSYELIHTQMKTQEWDSKSLGVQC EVQKQLKAFVTLERFDQLYGSTTSCQQAP KTKKFASSGSVFGKGVKFALKDGRVTTDII SVANEDGRRVAAILNHAHYLENLHFTIDG VDTHYFVKPGPSEGDLAILGLSGGRTLEN GVNVTVSQINTVLNGRTRRYTDIQLQYGA LCNTRYGTTLDEEKARVLELSRQRAVRQ AWAREQQRLREGEEGLRAWTEGEKQQVL STGRVQGYDGFFVISVEQYPELSDSANNIH FMRQSEMGR
2621	A	30	2298	LTRAPDPDRVGLVADFLRLFIPTAKGPVIN APLPQRLRSNTAPIRTLHAPSVHRPTGRES MPRTRLTRARTSPDTTGS DKTPHPRPKTLPI QTRSCADSGKLSEIRKIDDP LQHHLQNQSI QKSVKQCHEQNMFGNTVNQNGHFLKQ DCDTFDLHEKPLKSNLSFENQKRSSGLKNS AEFNRDGKSLFHANHKQFYTEMKFPALAK PINKSQFIKQQRTHNIENAHVCSECGKAFL KLSQFIDHQRVHTGEKPHVCSMCGKAFSR

456

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				KSRLMDHQRTHTELKHVECTECDKTFLLK SQLNIHQKTHMGGKPYTCSQCGKAFIKKC RLIYHHRHTHTGEKPHGCSVCGKAFSTKFSL TTHQKTHHTGEKPYICSECGKGFIKRRRLTA HHRHTHTGEKPFICNKCCKGFTLKNLSLTHQ QTHHTGEKLYTCSECGKGFSMKHCLMVHQ RTHHTGEKPYKNECGKGFKSPLIRHQRT HTGEKPYVCTECKRGFTMKSDLIVHQRT TAEKPYICNDCGKGFTVKSRLIVHQRTHTG EKPYVCGECGKGFPKIRLMGHQRTHTGE KPYICNECGKGFTKSHLNVHRRHTHTGEK YVCECGKGFTGKSMLIAHQRTHTGEKPYI CNECGKGFTMKSTLSIHQQTHHTGEKPYK NECDKTFRKKTCILQHQRFTGKTSFACTE CGKFSRLKNDLITHQRIHTGEKPYKCSDCG KAFTTKSGLNVHQKHTGERPYGSCDCGK AFAHLSILVKHKRIHR
2622	B	1	2034	MKLMETLNQCINAGHEMTKAIAAQFNDD SPEARKITRRWRIGEAADLVGVSSQAIRDA EKAGRLPHPDMEIRGRVEQVRVGYTIEQINH MRDVFGLRLRAEDVFPVIGVAAHKENT LLPFYLGEGKGDVYAIKPLAGRLTYFFLS GSARIENELMGKFVERKLATHTTSLDFWPL ETTPQLPPHILSPVFASASPSRCWRVASGK YCKVFRGSGFQAQXIPQPTLRDPHYVEDK GHKYLVEANTGTENGYQGEESLFNKAYY GGGTNFFRKESQKLQQAQKRDABELANGA LGHELNNDYTLKKVMKPLITSNTVTDEIER ANVFKMNGKWYLFDSRGSKMTIDGINSN DIYMLGYVSNSTGPKPLNKTGLVLQMG LDPNDVTFTYSHFAVPQAKGNNVNRFTQF RLSETKETNPYAMRLYESLCQYRKPDGSG IVSLKIDWILERYQLPQSYQRMPDFRRRFLQ GQFDHAASPVERGHLRKIPFRGGTRESRER GLSEAGYLPREAGQAQKRRPWTKGPLEKI GLETLHCDSRRYPCRSNWVWICTVKEGGR EGRGGRGRRVQLAAVAGTVAPAAAPKNP PPRFRWSVWARDGVKERVPLQAGVGGGQ AVQRRETARRSRGWLLRIWDSIGRDRSLG GNGFFTADQRFDAVLWLVAFRINSDKL
2623	A	513	796	TGTAWTPPPPLTTGAPCTPPPRCTARGRT/ PGDShLGGGAATAGGPRTSPMSSGGPSAP GMRPPASSPKRNTTSLNLSGLEPTFSFRITF GFM
2624	C	60	472	MPLEAYARNMLRTWSSLPWTRFRVCLLSL SLFLWANRLEDSCRSCQPNPMSLTTLPGHRL KEAVWLPAAPSRTMSPHLDPNQLGILLRVL KEKEDGDYPDMATHPSSRYEACSSGITL AAPPTHGPRPTDPRIGPAP
2625	A	1	1322	MAILPKVTYRFNAIPKLPVTFTELKTTLR FIWNQKRACIGKSVLSQKNKAGGITLPDFK

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LYYKATVTKTAWYWYQNRDIDQWNRTESEIMLHIYNHLIFDKPKDKKKWGKDSL FNKWCWENWLAICRKLKLDPFLTSHTKINSRWIKDLNVRPKTIKTLBENLGNTIQAIGMGKDFMTKTPKAMATKAKIDKWDLIKLSFCTAKETTIRLLGRPPALFTASSSVLKQLALEGILILD SRALLGFLYEARHSHSNPNHDAQNATSKKNIRDGYDKIYRQEQVLARMEETLITACCNVWCSHFRKQIGGQWLTLETKTTPQPFSSTSQISTDKDKGLNPQLKMDPGHMGWCPPGMGIPWQLSSDDRVWVLAAGSGRHPGSGFKSL/PGLLHEGSYGH****S*I*GGNS*GSSGGPQCISGEERVFRVVQSI
2626	A	129	329	VSNIVDPHQTVGLSTQEPGDIFTYSEFDGILGLAYPSLASE*SVPVLDNTMQRHLVAQDLFSVYMSR
2627	A	43	456	EFFHHVGDGLDLLTS*SAHLGLLKCDWYRREPPRPASDGHY*TDATGSLPSSGTT*IRTKPSQAPASWGLWNLAAHPPRSHPSCPMANLICSTLSSFDGGSPGTGPGGWCPGLSGSPARAVFKDSSCSLHPLATGI
2628	A	3	290	RQGFLCNHKGTVTADLQPLPPGLK*ISHL SLLSSWNYRCTPPHPADF*FFVERRSHYVA*ACLELLCSSLPALISQVRGITGMSTTPGPI CLL
2629	B	1	804	MVIVGLAAGVLLVGP GDGGLISEGVVREDLMCGVWSAGTWSVGTAERCLEKPGALHVIEGPLDSWDGPVMPNGPVKNHKGEEQVEPSKHPQMALEICLCLDFLYYFPLRGDASAGPVTWCTTSDTIILQQHRTLTSQGVDDFLKAKATFKASDFIDALVLSKDLNSGGRMELEIKCLIKVLELDLEGSSEPWKVLDKGVTVS YVFEMTIEGCLEGVNKSQETREGACGAGLEMAKEGSCLDERSSGTVSGYTQVSSELVCSGFLSPG
2630	A	322	549	GGGSSPRELAGAAGLTVTSQAVAARRQQPSFSRARAPAHSLRAALSLASSARSWGAVSRDRGPCPPAIMYQSSNKC
2631	B	1	384	MLVPVLILSPCLVGIEPWEVSPHTNSTSSYESTPKSYPLGTAAKAASGQSPSTTSPLPETAPSTLHERGLENVVCSDKDLRQATGYSAAEKSKPPGLCTRAFCPEAIPDAQDWVKCQPLGSL S ALNF
2632	A	1	275	KTSQDTKPSVLWKDVNSNLWCRPHDLLTWGRGYACVHIPSGPLGIPVQCKPYHGMA GTQCSTGNEECEPVGPAAPDNAASSDNTGPGWGM
2633	B	56	3476	XGKPEKFSFGLLDLPFRVGVFPNIPLEFQDEFGHTSQLVTDIQPVLEASGLSLHYEETNGPNCVIRGVTAKGPVNSCQGVAPNLPVYVVD C SSGTSILTGS AIQVQNIKKDQTLKARIEI



Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				<p>           PSCKDVPVEKTIKLLPSSHVARLQIFSVEG            QKAIQIKHQDEVNWIAGDIMHNLIFQMYD            EGEREINITSALAEEKIKGLLPDVQVPTSVKD            MRYCQVSFQDDHVSLESAFTVSMLELLQL            MVSLKTSNLLNNFRPLPDEPKHLKCEMKG            GKTVMGMQELQGEVVIITDQYGNQIAFS            PSSLSLSIAGVGLDSSNLKTTTFQSIPVINGR            DLQNPITVQLCDQWDNPAPVQHVKISLTKA            SNLKVKAITYNKSIEGPIIKLMILPDPEKPV            LNVKYDKDASFLAGGLFTGYVRPVPVPRS            LNSDISYFGVGGKQAVFFVQGSARMISKPA            DSQDVHELVLKEDFEKKEKNKEATYSGYI            RNRKISMFEKGKVPKIVNLREIQDDMQTLY            VNTAADSFEFKAHVEGDGVVEGIHPYHPFL            YDRETYDDPCFSPNNFGISFVHSLEVILXL            KDEDDEDDCFLEKAARGKRPIFCFWNGR            LIPYTSVEDRGLAPIECYNRISGALFTNDKF            QVSTNKLTFMDLELKLKDKNTLFRILNG            QEQRMKIDREFALWLKDCHEKYDKQIKFT            LFKGVITRPDLPSKKQGPWATYAAIEWDG            KIYKAGQLEPQALYDEVRTVPIAKLDRTV            AEKAVKKYVEDEMASLWILGYKPVQHMT            VLSTAGNCNTTFWKKINITVILRCRSLTKV            LLATERTFETAGVGGLILQVEEARLKEAQ            LRNELKIHNDIPTTQVPHIEALLKRKLSE            QEELKKKPRRSCTLPNYTKGSGDVLGKGQ            STGLGPVEVTQSSPSRTSEYFWLTKFCWL            EDWASGESLRLPLMVEGEGEPVYAEIHW            QKRDETVKDGVTLYLLQSVNQLLLTATKE            RIDFLPHYDTLVKSGMYEYYASEGONPLXI            YTHVGDREAQAALKLGRWSHPRTPNNAV            APGPPEGAGGGDAVTSQSALLTFSRTRFAS            GAHAGHPVLLRNEEEKGAPALVAPIFSAE            GPTCSLWWTLRPASTAGLKLPAARRVHATQ            PERAH         </p>
2634	B	1	384	<p>           MLASPLWLQALSAAAGTWRPRLGSGQAG            NSEMRAFLPGAGSQVRAQLQDRLPKTE            TKGALWPHTELCGMWSIAPGAENQELQID            SPLLGQLSNQVWREDGYGKAFLRLTLSSM            GITEEANENVLI         </p>
2635	A	628	1117	<p>           FFISVINGQVSSVQRLSGVGPACLSGGSANP            GPPPGTSPGAGAQR*PRADGSGSPQWPR            GARVGGGRLGTGGRGPGWRQVPRRLSP            GFGR*GGTGPGPVTSGKRGPSRRRAPAN            DKAACWPRFPGQPAS*TGFRGERGVKGFS            SWGSGWRAWEDGGTVH         </p>
2636	A	70	792	<p>           HGLVLDVRGPLSHAAPYWAPYPAATAAA            ARTAPLPPRSAIV*/SGPQPDFQELRKTWPS            QC/GMARREPLLPITAIPRVVVETTP*GFAK            QEPSVAGLRCRGSEAPA*LLHGVHRNVSE            TPGPEMGRPG*GNHRQRPKGQRGIPSSGLP         </p>

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GRCSGSRGPHSSPGQKPHGSTLSGRRGADP RPRRRVYLSTPLLCEKKPHHDITLKRKPGM GDGNNPCPWNAGLYGQATRFAPLPLCPRR RHGAVS
2637	A	571	172	SPLRPLLALALASVPCAQGACPASADLKH SDGTRTCAKLYDKSDPYENCCGGAELSL ESGADLPYLPNSWANTASSLVVAPRCELT VWSRQKGAKGTHKFSAGTYPRLEEYRRGI LGDWSNAISALYCRCS
2638	A	169	1144	INYSLEKHVVGALGRVLFSL* RERGLYLKHKRSGGGIW* SVSLTQGHHTVCSRSR* PHSVEVAGHVLVPATRAAVPCSASAGA* STYRTGVHQGNPTV* FLPSAVRGEPAKPLVDDLLPGWSLATHG QPPLVAAPGSLWGRPADA* CPRSTSRSPSGVQGCPLG* SEPPGSTSCCPRAP/T* WRSHGYLPPSREL* FARQRQGNRFDAAFESSGEDFHQMPRVGR MG
2639	A	1	1461	MRELYSIWLKGYWTEGDWAQSPPRSPREA LEGIRVHLRCFKAYGIIVLCQCPWNTPLLP VPKPGTKHYEPVQDLRLVNQATVTLHPTV PNPYTLLGLLPAEDTWFTHLDLKDAICSIRI APESQKLFAFQWEDLQSGVTTQYTWNWL PQGWVLKRVDALFQHLEDGKVPKKKS QICRQQVRYLGFTIWKGEHSLWSEKQVIC SLPEPKTRRQVREFLAGVGFCTLWIPNFAV LAKHLYGITKGGNWEFPEWGPLQQQAFLS ESPVEHNCVEVLDSVYSSRPDLRDQPWAS VDLELYLDGSSFNPQGERCAEYAVVTLDA VIETKPLPQGTSAQKAELIALTRALELSEGD CIWIKDCNIAPLRPRWKGPTVILTTPAV KRSIAIGNWQDEWLPERITQYYGPATWA QYGSWGYYNPIYMLNQMIWLQAVLEITTN KTGRALTILAWQETQMRNPTYQDRALDYL LLAEGGVCGKFNLTN
2640	A	254	418	MAISWKPTGLPWHSMQLQVLLAAWLPGPPTP TPHSALPSFSPPSLPPKMCLPKCC*
2641	A	433	3	ASFFNFSICICKIILEVGPPVGHPAHDDVGG RHGPGGR/GSRSPRSLQCAPGGRRSGCPA GSSPASTCPPSPGGSGADRFGPSPPPSREA APTAGAAASSTSSGASCPPVPASSRWGVR RTRSGSGGEREPRDRPSEPRLV
2642	A	2	798	VVEFADVEKKGAGRTEFRYPSYVQHIMGD IFSQGFPGFRWVCTSGDPQDLAVTDELATS VLEEAIADGVKVSVKLQYMDNIRWIREAA RHRLVVGSAQARILYSQKGRVAIAVAINQ AIACRIKAPVVLSDHHDVSGTDSFRET SNTYDGSACFADMAVQNFVGDACRGATW

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				VALHNGGGVGVWGEVINGGFLVLDGTPE AEGRARLMLSWDVSNGVARRCWSGNQK AYEHCQTMQENSTLVVTLPHKVEDERVLO QALQL
2643	A	1	2504	QISSGRELRVIQSEAGDAGLPRVEVILDCS DRQKTEGCRLQAGKECVDSPVEGGQSEAP PSLVSFVVSSEGTEQGEDPRSEKDHSRPHK HRARHARLRRSESLSEKQVKEAKSKCKSIA LLLTADPNPNSKGVLMFKRRRRARKYTL VSYGTGELEREADDEEEEGDKEDTCEVAFL GASESEVDEBLLSDVDDNTQVVNFDWDSG LV DIEKKLNRGDKMEMLPDITGKGALMF AKRRERMDQITAQKEEDKVGGTSPSREQDA AQTDGLRTTTSYQRKEEESVRTQSSVSKSY IEVSHGLGHVPQQNGFSGASETANIQRMVP MNRATKPPPGSVNQPATPFSPTRNMTSPIA DFPAPPPYSAVTPPDASFSGVSSPIAGPAQ PPPWPQAPWSQPAFYDSSERIASRDERISV PAKRTGILQEAKRRSTTKPMFTFKPKVSP NPELLSLLQNSEGKRG TGAGGDSGPEEDY LSLGAEACNFMQSSSAKQKTPPPVAPKPA\ VKSSSSQPVTPVSPVWSPGVAPTQPPAFPTS NPSKGTVSSIKIAQPSYPPARPASTLNVAG PFKGPQA AVASQNYTPKPTVSTPTVNAVQ PGAVGPSNELPGMSGRGAQLFAKRQSRME KYVVDSDTVQAHAARAQSPTPSLPASWKY SSNVRAPPPVA YNPIHSPSYPLAALKSQPSA AQPSKMGKKKGGKPLNALDVMKHQPYQL NASLFTFQPPDAKDGLPQKSSVKVNSALA MKQALPPRPVNAASPTNVQASSVYSVPAY TSPPSFFAEASSPVASPVVGIPTSPKQESA SSSYFVAPRPKFSAKKSGVTIQVWKPSVVE E
2644	A	938	652	RSSDGHAAETSRSCQLH*VSRSRNHPGPQP SGNTLRVRQSLSPDSRTLASAILAPP/TPLS SFRALALQPQEENRREEEMKEEGQVLGAV PLRTS
2645	B	182	394	MATHPSLLVCQVGLLGAQVPSVRAGMPQS RRQTEGAQGMVRNEEGSLRLSHHQACK ATHTQQWTLEVTAQ
2646	B	1	591	MTIHILLLLLLAFSAQGDLDTAARRGQH VPQHRGHVCYLGVCRTHLAEIYWIRCLH QGALGEGQPRAPGPLQLWAPPVARGGSPA RFPGRPAARGLAQCPARWVTSGTARPLL GFSLPIWLQRDMAEAHQAVGFRPSLTSDG AEVELSAPVLQEIYLSGLRSWKRHLRFW VRSGAGRFPSGDPGFCFRDV
2647	A	1	787	FQEAAVQLYSHAPHVQLRLKISPGHSPPAL GLSFPPGQGRGFSCQLLPASFWSGIPQRPLP QREPPGRTRTPAWSCSWGPAPPPVHTLVPA PSPGPGADRGGSGQPGLLVQGLPLGSLAP*

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				ALGLPGASADTPVPRRLHSQACCSHGVTG *GMG*GDVSPVPVPQGGLGWHLFRVPAGS QRSSPIPHQVLGGTRQPLGPGPVRKWTELA GDTGDKKEASSPKELVGPQRVGGLAGT VT LVPHLCCGRRAPPGGLDGAVEIVA
2648	A	2466	3395	KALCPCLPVPLVHGNVEVAGPRSGGACPT LGLVVLNPPGNHAATLRAHGQPCTALWR PLKPSQGYLEGAARGSAAKRPLQRALVS LDPGLGVLAATRLPGPVAGGWETQYMcC SAAAGSVGCQVAKQHVQDGRKERLEGFV KTFEKELSGDTHPGIYALDCEMSYYTYGLE LTRVTVDTDVHVVDYTFVKPDNEIVDYN TRFSGVTEADLADTSVTLRDVQAVLLSMF SADTILIGHSLESLLALKVIHSTVVDTSVL FPHRLGLPYKRSLRNLMA DYLRQIHDNDVD GHSSSEDAGACMHLVTWK
2649	A	178	556	QSPQEHFHPECGRRDILCQVRQEIRWPNPG EVVHHLGLEICPVWILQLHLALRTAPEHPL QVHRPGGGAV*RGVPPPLRLQACDGPEV PAAGRPRPARSSPGQWPP*/PAAVAPPVTE RPPTPSAA
2650	A	803	1068	RAMEPLLLGRGLIVYLMFLLKFSKAIEIPS SGKVKTFSAILLSMDSFPQAGGIFGT PPGLG SRILSPSPMVSLGSCCTHRSPICFSP
2651	B	1	559	MAERAAGGQLPSQGPVQLPSTRKEKDEQT ENQQLFFIRQRTESPGKARPPNLETQTS GFQ EPQLTGAEPLRGQCHGLEPLMNFWRCHL DKTNLRLKEELKAEKKSGFWDNLVLKQNI QSKKPDEIEGWEPKLALEDISADPEDTVG GHPSWSGWEDDAKGSTKYTSLASSANS SR WSLRAAGKAX
2652	A	1	526	FRLGRKPR*GGVM*PVWSRGEPSVGA EA G/RS*SAPRRLHHPAAGLATGLSASGRRS ARWKMERASGLSPGGGLGATSRQMSPGT QLANPPDHGDKDCLGRISPGSGKQIQAAG QLPGPPTSLAPAQGRRLRSLTPWGLQTP EHS EPEGIGHLQAATEAVLPHSTQNLITKRNL M
2653	A	3	396	AA YTL LLLHAE LLQWSDKPCVP HLLQRDSY YVYTQQLKEKLYQEIISYFDKGKMW EKA IKLSKELAETYESKVFDYEG LGNLLKKRAS FYENIKAMSPQPEYFAVGYYGQGFP SF LR NKIFIYRGKEYER
2654	C	1	507	MPLTHPNHGPD TLQRWTSSQTPTSLSSKLN PEPEADAASILIATSILYKQSDPYLDILARV YGPPTAAEENLKCLKEQGQAHLRHFL LCK MAPIAVVLTAAMFENWTHRRQWQVFEP GAREEEKSLKSPRFLALKVLRKGAD FQRL RLYQANMGQAKLPLALFHPLC
2655	A	178	1206	ALMNKCAVSTGRQRCSVMWARACSVFCV LTLRNTGAQKHWLTEGAAKEHCVSDDSE HFESWRAAQLFESVDAEPMNMESQLHFIM

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				PKALRTKKAASDSSKEQVANSRESSPSPKE VNDSPRAATKSPESQNLIDGTTKPSLKQPD SPRNISSDNSSKGTTPSSPAGSTTAIPKVRIT IKTSSGEIKRTETRVFPEVDLDSGKKPSEQM VSVMASVTSLSSPASAAALSSPPRVPLQS AVVTNAVFPAPETPKQVTIKPVATAFLPVS AVKTAGSQVINLKLANNNTTVKATVIPAAS VQSASSIIKAANAIQEQAVMMPASSLANA KLVPKTVHLANLNLLA
2656	A	215	389	KGAGVLQTFGSSSEVFCIDVDRELLIFAYQ NILLFLKNKRALILETTTCFGWVGTVKRT
2657	A	1	737	FRGEIAENLPEQDILIQSV CETMVPKLV AED IPLFLSLLSDVFPVGVQYHRGEMTALREELK KVCQEMYLT YGDGEEVGGMWVEKVLQL YQITQINHGLMMVGP SGSGKSMAWRVLL KALERLEGVEGVAHIIDPKAISKDHL YGTL DPNTREWTDGLFTHVLRKIIDSVRGELQKR QWIVFDGDVDPEWVENLNSVLD DNKLLTL PNGERLSLPPNVRIMFEVQDLKYATLATVS RCGMVWFSED
2658	B	41	166	MKIAALLGCMMAARCGTLSAMRDL SFS DENRR LAVGTA AAA
2659	A	1	894	MPGPMSLWLLLLVPLSLEHSDLRICFP GQ VVSMESSSTGFIWTDVRAWQTSNRHVSSW REPRHSRMPPGAGLMERIQAI AQNVSDIAV KVDQILRHSLLH SKVSEGRRDQCEAPSDP KFPDCSGKVEWMRARWTS DPCYAF FGVD GTECSFLIYLSEVEWF CPLL PWRNQTA AQR APKPLPKVQAVFRSNLSHLLDLMGSGKES LIFMKKRTKRLTAQWALAAQRLAQLGA TQRDQKQILVHIGFLT EESGDVFSRVLKG GPLGEMVQWADILTALYVLGHGLRVTVSL KELQR
2660	A	3	14703	AAAVSARRAAAGGSRGAGGWGTADASG AMAEGGEGGEDEIQFLRTEDEVVLQCIATI HKEQRKFCLAAEGLGNRLCFLEPTSEAKYI PPDLCVCNFVLEQSLSVRALQEMLAN TGE NGGEGAAQGGGHRTLLYGHAVLLRHSFS GMYLTCLTTSRSQTDKLA FDVGLREHATG EACWWTIHPASKQRSEGEKVRIGDDL LVS VSSERYLHLSVSNQNIQVDASFMQTLWNV HPTCSGSSIEEGYLLGGHVVR LFGHDECL TIPSTDQND SQHRRIFYEAGGAGTRARSLW RVEPLRISWGSNIRWGQAFRLRHLTTGHY LALTEDQGLILQDRAKSDTKSTAFSFRASK ELKEKLDSSHKRDI EGMGVPEIKYGDSVCF VQHIASGLWV TYKAQDAKTSRLGPLKRKV ILHQEGHMDDGLTLQRCQREESQAARIIRN TTALFSQFVSGNNRTAAPITLPIEEVLQTLQ DLIAYFPPEEEMRHEDKQNKLRSLKNRQ NLFKEEGMLALVLNCIDRLNVYNSVAHFA

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GIAREESGMAWKEILNLLYKLLAALIRGNR NNCAQFSNNLDWLISKLDRLLESSSGILEVL HCILTESPEALNLIAEGHIKSIIISLLDKHGRN HKVLDILCSLCLCNGVAVRANQNLICDNL LPRRNLLLQTRLINDVTSIRPNIFLGVAEGS AQYKKWYFELIIDQVDPFLTAEPHTLRVG WASSSGYAPYPGGGEGWGGNGVGDDLYS YGFDGLHLWSGRIPRAVASINQHLLRSDD VGKLLPGPRGCPASHSASMGSPCRGCLENF NTDGLFFPVMSFSAGVKVRFMLGGRHGEF KFLPPSGYAPCYEALLPKKMRLEPVKEY KRADAGIRDLLGTTQFLSQASFPCPVDTSQ VILPPHLEKIRDRLAENIHFWGMNKIELG WTFGKIRDDNKRQHPCLVEFSKLPETKKN YNLQMSTETLKTLLTLGCHIAHVNPAAEE DLKKVKLPKNYMSNGYKPAPLDLSVDK LLPPQEILVDKLAENAHNVWAKDRIKQGW TYGIQQDLKNKRNPRVLPYALLDERTKKS NRDSLREAVRTFVGYGYNIEPSDQELADSA VEKVSIDKIRFFRVERSYPRSGKWYFEFE VVTGGDMRVGWARPGCRPDVELGADDQ AFVFEENRGQRWHQSGGYFGRTWQPGDV VGCMINLDDASMIFTLNGELLITNKGSELA FADYEIENGFPICCLGLSQIGRMNLGTD STFKFYTMCGLQEGFEPFAVNMNRDVAM WFSKRLPTFVNVPKDHPHIEVMRIDGTMD SPPCLKVTHKTFTGQNSNADMIYCRLSMP VECSSFSHSPCLDSEAFQKRKQMQEILSH TTTQCYAIRIFGGQDPSCVWVGWVTPDY HLYSEKFDLNKNCTVTVTGLGDERGRVHES VKRSNCYMWVGGDIVASSQRSNRSNVDL EIGCLVDLAMGMLSFSANGKELGTCYQVE PNTKVFPVFLQPTSTSLFQFELGKLKNAM PLSAAIFRSEENPVQPQPRLDVQTIQPV WSRMPNSFLKVETERVSRHGWVVCLEP LQMMALHIPEENRCVDILELCEQEDLMRF HYHTLRLYSAVCALGNSRVAYALCSHVDL SQLFYAIDNKYLPGLLRSGFYDLLISIHLS AKERKLMKNEYIPIITSTTRNICLFPDESK RHGLPGVGLRTCLKPGFRFSTPCFVVTGED HQKQSPEIPLESRLTKALSMLTEAVQCSGA HIRDPVGGSVFQFVPVCLKLIGTLLVMGVF DDDDVRQILLIDPSVFGEHSAGTEEGAEK EEVTQVEEKAVEAGEKAGKEAPVKGLLQT RLPESVKLQMCELLSYLDCDELQHRVEAIV AFGDIYVSKLQANQKFRYNELMQALNMS AALTARKTKEFRSPPEQINMLLNFLQGEN CPCPEEIREELYDFHEDLLHCGVPLEEEEE EEEDTSWTGKLALVYKIKGPPKPEKEQPT EEEERCPTTLKELISQTMICWAQEDQIQDSE LVRMMFNLLRRQYDSIGELLQALRKTYTIS

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				HTSVSDTTNLLAALGQIRSLLSVRMGKEEE LLMINGLGDIMNNKVIFYQHPNLMRVLGM HETVMEVMVNVLGTEKSQIAFPKMVASCC RFLCYFCRISRQNKAMFEHLSYLLNSSV GLASPSMRGSTPLDVAASSVMDNNELALS LEEPDLEKVVITYLAGCGLQSCPMMLAKGY PDVGWNPIEGERYLSFLRFAVFNSESVEE NASVVVKLLIRPECFGPALRGEGGNLLA AMQGAIKISENPALDPSQGYKREVSTEDD EEEEIVHMGNAIMSFYSALIDLLGRCAPE MHLIQTGKGEAIRSILRSLVPTEDLVGIISI PLKLP SLNKDGSVSEPD MAGNFCPDHKAP MVLFLDRVYGIKDQTFLLHLLVGFPLDLR ASASLDTVSLSTTEAALALNRYICSAVLPL LTRCAPLFGGTEHCTSLIDSTLQTTYRLSKG RSLTKAQRDTTEECCLAICNHLRPSMLQQL LRLVFDVPQLNEYCKMPLKLLTNHYEQC WKYYCLPSGWGSYGLAVEEELHLTEKLF WGHDLSLHKKYDPLFRMALPCLSAIAGA LPPDYLD SRITATLEKQISVDADGNFDPKPI NTMNFSLPEKLEYIVTKYAEHSHDKWACD KSQSGWKYGISLDENVKTHPLIRPFKTLTE KEKEIYRWPARESLKTM LAVGWTVERTKE GEALVQQRENEKLRSVSQANQGN SYSPAP LDLSNVVLSRELQGMVEVVAENYHNIWA KKKKLELESKGGGSHPLLVPYDTLTAKK FKDREKAQDLFKFLQVNGIIVSRGMKDME LDASSMEKRFGYKFLKKILKYVDSAQEFIA HLEAIVSSGKTEKSPRDQEI KFAKVLLPLV DQYFTSHCLYFLSSPLKPLSSSGYASHKEK EMVAGLFCCLAALVRHRISLFGSDSTTMV SCLHILAQTLDTRTVMKSGSELVKAGLRAF FENAAEDLEKTSENKLGKFTHSRTQIKGV SQNINYTTVALLPILTSIFEHVTQH QFGMDL LLGDVQISCYHILCSLYSLGTGKNIVVERQ RPALGECLASLAAAIPVAFLEPTLNRYNPL SVFNTKTPRERSILGMPD TVEDMCPDIPQL EGLMKEINDLAESGARYTEMPHVIEVILPM LCNYLSYWWERGPENLPSTGPCCTKVTS EHL SLILGNILKIINNNLGIDEASWMKRIAV YAQPIISKARPDLLRSHFIPTLEKLKKKAVK TVQEEELKADGKGDTQEAELLILDEFV LCRDLYAFYPMILIRYVDNNRSNWLKSPDA DSDQLFRMVAEVFILWCKSHNFKREEQNF VIQNEINNLAFLTGDSKSKMSKAMQVKS GQDQERKKTKRRGDLYSIQTS LIVAALKK MLPIGLNMCTPGDQELISLAKSRYSHRDT EEVREHLRNNLHLQEKSDDPVKWQLNL YKDVLKSEEPFNPEKTVERVQRISA AVFHL EQVEQPLRSKKA VWHKLLSKQRKRAVVA CFRMAPLYNLPRHRSINLFLHGYQRFWIET

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				EEYSFEEKLVQDLAKSPKVEEEEEETEKKQ PDPLHQIILYFSRNALTERSKLEDDPLYTSY SSMMAKSCQSGEDEEEDEDEKKTFEEM EKQKTLYQQARLHERGAAEMVLQMISAS KGEMSPMVVETLKLGLAILNGGNAGVQK MLDYLKEKKDAGFFQSLSGLMQSCSVLDL NAFERQNKAEGGLMVTEEGTLVRRERGEK VLQNDFTRLDRLFLQLLCEGHNSDFQNFL RTQMGNNTTVNVIISTVDYLLRLQESISDFY WYYSGKDIIDESGQHNFASKALAVTKQIFNS LTEYIQGPCIGNQQSLAHSRLWDAVVGFL HVFANMQMKLSQDSSQIELLKELLDLLQD MVMMLLSLLEGNVNGTIGKQMVDTLVE SSTNVEMILKFFDMFLKLDLTSSDTFKEY DPDGKGIISKKEFKAMEGQKQYTQSEIDF LLSCAEADENDMFNYVDFVDRFHEPAKDI GFNVAVLLTNLSEHMPNDSRLKCLLDPAE SVLNYFEPYLGRIEMGGAKKIERVYFEISE SSRTQWEKPQVKESKRQFIFDVVNEGGEQ EKMELFVNFCEDTIFEMQLASQISESDSAD RPEEEEEDEDEDSSYVLEIAGEEEDGSLEPAS AFAMACASVKRNVTDFLKRATLKNLRKQ YRNVKMTAKELVKVLFSSFFWMLFVGLF QLLFTILGGIFQILWSTVFGGGLVEGAKNIR VTKILGDMPTDPTQFGIHDDTMEAERAEM EPGITTEL VHFKEGKGD TDIMSDLFGLHPK KEGSLKHGPEVGLGDLSEIIGKDEPPTLEST VQKKRKAQAAEMKAANEAEKVESEKAD MEDGEKEDKDKKEEQAEYLWTEVTKKKK RRCGQKVEKPEAFTANFFKGLEIYQTKLLH YLARNFYNLRLALFVAFAINFILLFYKVTE EPLBEETEDVANLWNSFNDEEEEMVFF VLQESTGYMAPTLRALAIHTIISLVCVVG YCLKVPLVVFKEKEIARKLEFDGLYITEQ PSEDDIKQWDRLVINTPSFPNNYWDKRV KRKVINKYGDLYGAERIAELLGLDKNALD FSPVEETKAEASLVSWLSSIDMKYHIWKL GVVFTDNSFLYLAWYTTMSVLGHYNNFFF AAHLLDIAMGFKTLRILSSVTHNGKQLVL TVGLLAVVVYLYTVVAFNFRKFYKNSD DDEPDMKCDDMMTCYLFHMYVGVRAGG GIGDEIEDPAGDPYEMYRIVFDITFFFVIVI LLAIHQGLIIDAFGELRDQEQVREDMETK CFICGIGNDYFDTPHGFETHLQEHNLAN YLFFLMYLINKDETEHTQESYVWKMYQE RCWDFFPAGDCFRKQYEDQLG
2661	C	54	350	MLNSSEQRPHGVLDVWPVGIHGALCAGR WLRTGQLSWDTRHMLARKMVSSSEPQR PTSWSWCLLASTVRPLLVDGSGWGSCRGR PAACWKEDGQFF
2662	A	50	646	SSALLSSNQTA SFGSCSLSLPCSARERTPEG



Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GGWPGGRLSEPLPAMLLLVVSVVAALAL AVLAPGAGEQRRRAAKAPNVVLVVSDSFD GRLTFHPGSQVVKLPFINFMKTRGTSFLNA YTNSPICCPSRAAMWSGLFTHLTESWNNF KGLDPNYTTWMDVMERHGYRTQKFGKL DYTSGHHSIRHSERGSTNQSEK V
2663	B	44	293	MPVWRRRRRLRARSWALRARPLSLPRAQ RSGRLLRRPKGYAPGAPKAHELSPQAICAV AFX
2664	C	40	495	MVILNALQRR AFLCAANVKIPRLRIKVKTK EASAQVVKEECNKYLLFLLPVPSAGLLPSI MEIADPFSSFGSEDKCYTLTPPLPRHTEKSS DSQEKGHFEAGVEPKSRGSTPGQYPGIGCF ARFREYQIGMRHLTTRPAMHRAQVLFPLS F
2665	A	587	2	FLTRETGDPTGRSSSHANTQSRFFDPPG\ PLNNLGNTHGCGRRAGRCPGTGPDPG\AG CGGPRCWPSGHLAATGD*GPSCGRLGANR GEAGPAGFTACSPLSGCRTPTYTHHFPASRM SCHLNCASPRTYRSQGNRGCEVAQGSQG AGGERGAKSQVPVPAPARNKDPKCRKPR NRRPGNSGPVVRA YRRQR
2666	A	1	1853	RARRLALQCHVCVCALTPGEQSGRRLLPGQ TWLMFSCFCFSLQDNSFSSTTVTECEDPV SLHEDQTDCCSLRDENNKENYPDAGALVE EHAPPSWEPQQQNVEATVLVDSVLRPSMG NFKSRKPKSIFKAESGRSHGESQETEHVVS SQSECQVRAGTPAHESPQNNAFKCCQETVR L\QPRIDQRTATSPKDAFETR\QDLNEEEAA QVHGVKDPAPASTQSVLA\DGTDSDPSPV HKDQGNEADSAPEDLHSVGTSRLLL\YHIT DGDNPTAVRHGCSL\FSGQSQRFNLDPEA PSPPSTQQFMMPRSSRCSCGDGKEPQTIT QLTKHIQSLKRKIRKFEEKFEQEKKYRPSH GDKTSNPEVLKWMNDLAKGRKQLKELKL KLSEEQGSAPKGPPRNLLCEQPTVPRESGK PEAAGPEPSSSGEETPDAAALTCLKERREQL PPQEDSKVTKQDKNLKPLYDRYRIKQILS TPSLIPTIQEEEDSDEDRPQGSQQPSLADPA SHLPVGDHLTYSNETEPVRALLPDEKKEV KPPALSMNLHEATMPVLLDHLRETRADK KRLRKALREFEEQFFKQTGRSPQKEDRPM ADEYYEYKHIKAKLRLAPPAGSYFP
2667	C	147	398	MYKAQFLAASPGRC LGLLAASNHHAKSIH GFRRLVKTMNRRLCSLCQFPPLPKHLLSL WFGDQGHTSQYFTLSTQRNEAQLQ
2668	A	1	1787	MSKGESRKCN EENVKSSKVVKVFIVLTPQ FLSRDKDQLTKELQKHVKS VTVSCKSPRK LLSHITRLHPPSKGQGENLTHLVDSIKATTW CQPVWETVEGQRRRVGNCIDFTNGCDLVG SSSLHNMLVCSSYDINRQDTFQKDR TSEKH

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LLDSVFTALQDSAGQQWPAPLHPQRGEEV ADPRGAPSRHVEPENSSPCQGNGEQAGKA GARALCGQARRSPATMPPPLTTRSLCEFAV FLHHLFPELFHYRKLGEQDSCYGDGGKQ ELDPQRLQIICNFTEVYFPHMQEEAWRQA GPGPAEAAAD/TATSRRSTSPTRRRRPGCS GAPSASTTSFRAWGWTOAAKASPPRDNCY NSSSLPDDISLFTHDNLHKQHSCSDSLGKK QLDPSCIKLIRH*VHLLYLCTKNRNVWTL FMGNLHWNRRNRGAPTSSSARSTCWPRV*R HEELCNQS*EVQRGV*GSPAAPERSSKDFC KIPLDEVVVP*/DFPVRSPYLLSDKEVCKI VQQSLSVGNFAAGLL/LPRTSSCSTTIFGL/ DNKKQLDPTQLRLICH*VEAVYPVEKVEE VWHCECIPSNDEQCHCPNRKKCNILKKA KVEK
2669	A	14	425	RRFREPDQAQMLEIPNLTPYTHYRFRMKQV NIVGSPSPSSRVIQTLQAPPDVAPTSVT RTASETSLRLRWVPLPDSQYNGNPESVGY RIKYWRSDLOSSAVAQVVSRLEREFTIEE LBEWMEYELQMQAFNAV
2670	B	1	825	MRALKLQRRKSFVIVVAWEAFVQLVNYE CKVGEWKGLAHCVSQNNKYRTTYIAGVP NPQEPGYTAGGQLKGNLTVLHLLVIEGK WEAVRKFPFKYIVNTAIVKEARKYWVEE GSSLAKATRSNPGYLQPYMRTGIPVFAPPK LPFGPPCPLSCTHINPKPQAPADQQLPIHL AESHFHHSIKPRIHPSSPCVTRFFLDAEREL GIQKAVPWSFTLVKKQKSLGLPSVQDFGS VYKMNWSDVACCDPQLQQAASAQTSAT SQLSRVTES
2671	B	475	848	XRTERVHLRITPGDDSRKRSSASHYRVA SRLTSLDREQLYLEQSTEGPEQDKREGKS ARSSSREPTGQPTLLGGMRRARKRTLVL GPFPRVISGSNAKMDTLSPACACAFALYGI PKPAA
2672	A	3	765	LGTVSYGADTMDEIQSHVRDSYSQMOSQA GGNNTGSTPLRKAQSSAPKVRKSVSSRIHE AVKAIVLCHNVTPVYESRAGVTEETFAE ADQDFSDENRITYQASSPDEVALVQWTESV GLTLVSRDLTSMQLKTPSGQVLSFCILQLFP FTSESKRMGVIVRDESTAEITFYMKGADV MSPIVQYNDWLEBECGNMAREGLRTLTV AKKALTEEYQDFEVSRLPGIPSSY/DRCLP YAEISSSCLCMKLELGSL
2673	A	9	413	EPKSLIQIKQSIVELKLAEDSFVLKVVQL EELLQVRHVSFIVGNAGSGKSQVLTASNE RIPLNRTMRLVFEISHLRTATPATVSRAGIL YINPADLGWNPVSSWIERRKVQSEKANL MILFDKYLPTCLDK
2674	A	379	17	SWGVMYKYQPLDLVRRYFGEKIGLYFAW

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LGWYTGMLFPAAFIGLFVFLYGVTTVDHS QVSKEVCQATDIIMCPVCDKYCPFMRLSDS CVYAKVTHLFDNGATVFSAVFMAVWATV LMEFGK
2675	A	1	1833	MVDSLIRVGV MARGNAITLPVCGRDVKF TLEVLRGDSVEKTSRVWSGNERDQELLTE DALDDLIPSFLLTGQQTAFGRRVSGVIEIA DGSRRRKAALTESDYRVLVGELDDEQM AALSRLGNDYRPTSAYERGQRYASRLQNE FAGNISALADAENISHSDKFDANDPILKDQ TQEWSGSATFTSDGKIRLFYTDYSGKHYG KQSLTTAQVNVSKSDDTLKINGVEDHKTIF DGDGKTYQNVQQFIDEGNYTSGDNHTLRD PHYVEDKGHKYLVEANTGTENGYQGEES LFNKAYYGGGTNFFRKESQKLQQSACKRD AELANGALVNTQSTTTTRPGSNSLSHLMW PVDHQKFQSVTEMCGSILSRDFADFGTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLL PVSLVKRKTTLAPNTQTASPRALADSLMQ LARQVSRLESGQDFADFGTTIKQDFRLLGQ TSVDRLLQLSQGQAVKGNQLLPVSLVKRK TTLAPNTQTASPRALADSLMQLARQVSRL ESGQDFADFGTTIKQDFRLLGQTSVDRLLQ LSQGQAVKGNQLLPVSLVKRKTTLAPNTQ TASPRALADSLMQLARQVSRLESGQ
2676	B	1	309	MGKAMLQLLIRAHWTVFPCEHEDNAASV SVTLCSDLAGGEVVSAVLTGQSVVQTEKEI DRSSKPPACLVAPQVVFCEVLRVDESYHR KYPVQLRPVHIAAK
2677	A	2	179	RGKKSVTTVAGPMAQDVESLALCLQALLS EDMYRLDPTVLQMPFREEVKTPFPTPGCSE
2678	A	34	390	MKRRRQLRARVFALALAWSLGPCWALRV AVPKASXTIRGPQRRLLASLLQENTEILGY LLGSVAAFGSWASRIPPLSRICRGKTFPSIH LWTRLLSALAGLLYASAIAAHDRHPEYLL R
2679	A	568	3	SYYERINRQLIEAKMALQDREKMEKVFD DIETNMNLIGATAVEDKLQDQAAETIEALH AAGLKVVVLTGDKMETAKSTCYACRLFQ TNTLELTLTKTIEESERKEDRLHELLIEYR KKLLHEFPKSTRSFKKAWTEHQEYGLIIDG STLSLILNSSQDSSSNYKSIFLQICMKCTA VLCCRMAPL
2680	A	3	394	SSRWAFQVLSPSADSARLPGRAPGDRDCTF QPSAPAPSKPFLSTPPFYSAACCGGSCRRPA SSTAFFREESMLPLLTQDSNSKARRGILRR AVFSEDQRKALEKMFHKQKYISKTDKCKL AINLGLKESQ
2681	A	42	406	EPGDPREGEEEEDEPDPEAPENGSLPRFV PRFNFSCLKDLTRFVDFNIKGRDVIVFLHIQK TGGTAFGRHLVKNIRLEQPCSCKAGQKKC

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				TCHRP GK KETWLF SR FSTGW SCGLHADWT E
2682	A	10	932	LQLCSMWLLRSWVQAEGAVSISDSPFSLH QCWAVLHKA WCVFLQLPGGFTFTLNPLSD NLLGKRVD SAPSWGPLGSAFRGVHMP CV GAAWEGKGP NLLRPSGKLGPSGSRPTPIGQ QQLPEVPRAKGPLGPAAVICQ/HMPAPSTG GKRGSFSGRYLSASLELGGLPMAPTGPSAL SAPPSVSRGAR*STREKPGVYASAT*AAEIR EGQALGGVPRPSRNG/SGGPLGPDGPNPGK LRRSKAGCPWWHLSSVDAGE*LWKQHST AVFSMPGTQPPWRGLITMPISPRGTEPTAH PGPRSPGLAYSLTA
2683	A	1	416	NRLTTHSPHSPGPGGRQAPWRRQCRPASC PAKSTTWPVTRAPTRPPAWPPASAPP/R Y LLEEFQNCYARYHQAFADRDQSERQRH ESQQLATETQALAQRTQQDSTRTVGERLQ DTHSWKSELQREMEALAAETNLLL
2684	A	356	1356	TPTTSGRTRKMWPRPGT*PP/ANCSANINLT HQPWFQVLEPQFRQFLFYRHCRYFPMLLN HPEKCRGDVYLLVVVKS VITQHDRREAIR QTWARAAVRGWGPSAVRTLFLLTASKQ EERTHYQQLLAYEDALYGDILQWGFLDTF FNLTLEIHF LKWLDIYCPHVPFIFKGDDD VFNPTNLLEFLADRQPQENLFGDVLQH ARPIRRKDNKYYPGALYGKASYPPYAGG GGFLMAGSLARRLHHACDTLELYPIDDF LGMCLEVLGVQPTAHEGFKTFGISRNRNSR MNKEPCFFRAMLVVHKLLPELLAMWGL VHSNLTCSRKLQVL
2685	A	1	741	VRSMSCPPSWPYCAPCPTNIGESTSPLR KTI ETPTLWDPKAPSCSLELPWVLAS PQRSRG TALPFLPSNVLP SLALPSTSF LCRPLL SHLV TSLLAGPGAHDGHLRKEGWRSTPEMTSLP APEHPASPCDSVLCSPDVSMCTLGPAARW DAQAKSAPLPPCCTDCKSEPHLQRPWAQP HTSQATSVDSGEAGTKGMSQFTVWTTWWR SRPCETRQGE GIGNWGYSVTPGPPGSQNL P ARLDGQGLAS
2686	A	396	687	TFCPRCGCP SGLAMRLFLSLPVLVVLSIV LEGPAPA*GAPEVSNPFDGLEELGKLTLEDY TREFINRITQSELP AKMWDWFSETRKVK E KLKTDS
2687	A	2	3794	PRGPRPGASGSAMWLSPEEVLVANALWVT BRANPFFVLR RRRRGHGRGGGLTG LLVGTL DVVLDSSARVAPYRILHQTQDSQVYWTVA CGSSRKEITKHWEWLENNLLQ TLSIFDSEE DITTFVKGIHGIIAEENKNLQPQGD EDPG KFKEAELKMRKQFGMPEGEKLVNYYSCS YWKGRVPRQGWLYLTVNHLCFYSFLLGK EVSLVVQWVDITRLEKNATLLFPESIRVDT

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				RDQELFFSMFLNIGETFKLMEQLANLAMR QLDSEGFLEDKALPRPIRPHRNISALKRDL DARAKNECYRATFRLPRDERLDGHTSCTL WTPFNKLHIPGQMFISNNYICFASKBEDAC HLIPLREVTIVEKADSSSVLPSPLSISTKSK MTFLFANLKDRDFLVQRISDFLQKTPSKQP GSIGSRKASVVDPTSESPAPQEGSEQPASP ASPLSSRQSFCAQEAPTASQGLLKLQKNS PMEDLGAKGAKEKMKESWHIHFHEYGR GVCMYRTAKTRALVLKGIPESLRGELWLL FSGAWNEMVTHPGYYAELVEKSTGKYSL ATEEIERDLHRSMPEHPAFQNELGIAALRR VLTAYAFRNPTIGYQCAMNIVTSVLLLYGS EEFAFWLLVALCERMLPDYYNTRVVGAL VDQGIFEELTRDFLPQLSEKMQDLGVISSIS LSWFLTLFLSVMFPFSAVVIVDCFFYEGIK VILQVALAVLDANMEHLLGCSDEGEAMT MLGRYLDNVVNKQSVSPPIHLRALLSSSD DPPAEVDIFELLKVSYEKFSSLRAEDIEQMR FKQRLKVIQSLED TAKRSVVRAIPVDIGFSI EELEDLYMVFKAKHLASQYWGCSTRMAG RRDPSLPYLEQYRIDASQFRELFASTPWA CGSHTPLLAGRMFRLLDENKDSLINFKEFV TGMSGMYHGDLTEKLKVLKYLHLPALSP EAEAE\SALEATHLFSQRDSSSEASPLASDL LFLPWAEALPQEEQEGSGSEERGEEKGT SSPDYRHYLRMWAKEKEAQKETKDLPK MNQEQFIELCKTLYNMFSEDPMEQDLYHA IATVASLLLRIGEVGKKFSARTGRKPRDCA TEEDEPPAPELHQDAARELQPPAAGDPQA KAGGDTHLGKAPQESQVVVEGGSGEGQG SPSQLLSDDDETKDDMSMSSYSVVSTGSLQC EDLADDTVLVGGEACSP TARIGGTVDTDW CISFEQILASILTESVLVNFYEKRVDIGLKIK DQKKVERQFSTASDHEQPGVSG
2688	B	119	682	GDKGADEREISGGTDTAAAAQLKIHYP GPSTVQEHKEVFNTKLADGQNGSPSKQASI CDRQFVVAGGYHRS LADEAYGDEEDLPK VVGLVHSTRGPAHPTYLLRPLQKDQDSSL LRASGGGGGSPSSSTKSEHSCRQIHIPGPF HADITGQKWFPGGVSTEPARNMGFLKPTP TPLLRS PKDFR
2689	B	1	3097	MAGARVGPAAGARTAVPAAGEVPASPAL TDTQKGTGIGHWVVA VAPTQTSVWPKPF RGNRISVLGFEPHSLVSADPQQSQYPYFLFP EPPSPKPLSMLEDSYASLKIQASARAPPLSPI DMDKQERIKAEKRRLRNRIAASQVPQAQA GAHLAPGKKVKTLKSQNTELASTAACAS SSSLVGGSRRERSSESGPHICAQRAPPRRAL ARGRLMPGDTGPRELHRNPSVVVVVCLLV SLLIGSVMAVRFC HRNESKFENLDEVS

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				MGSVNDRLSFAHHLQEHQFLFPRVAGCRA RGTPTPAALGRCWPWPLRPPCPASQRQK VAVGPKRMGSPFRLAATVRQPERPQAPM AVPSCPSTPDYENMFVASQQPSTSGMNGK KALPAGILQMVTDTSRPNVGGDESLDCLV LNRISYTCRSTLSRPSFSAPGREESGSVMA PDDSMGIMRSLGGLSRLTVAIVRDVTKFC DPGPPHPALQETPQMAPSPGAPQLNPPAP PRKRNTASAPVHLRAARDDEAALYPFLQ VSYSLSGHKNNTYTYAWVVGGFRALGYK HSTDVCSGVTIQEEMWIRHRLRAAPISQR TRHYHRLGCSMAGSGCASDLLCCDWRD SCCRCSLSAAQATPLSSPRPRPSAARLSAR GAATTAGSVCSGGGEVAGEPGPRRHVHG GAEKWGDVQWTPGDCDNWMNINLREVIC TSGTGQVLADARVLHPRQHHQYLRIPEII DMVKEEVGPRAAAEACSSRSSRKPRHGR RWPRCFGALSCCGGRESSTCKPLPFADP QVLHAPEKGVWEAGSRTRPRERAPRSVCP GSGPGPGVEATARSCRAGGAEEVEGGTGA QASMVNTTGYWARPLQATQGGSAWQQ WGTREASPDTTTTRGLTGAKPESTNSQNH
2690	A	1007	537	SRKGSSLAHPLSPSRLSAVPTAGGGGDSE AKPHLVSPGGSEGATWCGHGGQGRGGSGND RGGQ\GPGAGGRRGIPTPARGAVITYKTQRR EEEGTRGCNQLASLSGPQGATVSPSSGGSS PGTCCDRHPLRADTRMMVWGQEPSPLVC FPKLQPDLSL
2691	A	1	1656	METEPSKAKANDPGSAAEGVVFASISSGLG EVTFLSLTAFYPRAVISWWSSGTGGAGLLG ALSYLGLTQAGLSQQTLTLLSMLGIPALLA SILRKALDKIAEIKSLEERRIGHKYLGRLY CPPLYVL YTDAFWSVTPYSEVHIAFTILEEV SLCDSKLIHIFVRLAYACPRFTVSAWAASI PEYMVRIISLLTAQVDMTIIGIAFMPCPRPL MPTVAPTAAREMGVHHTGDSAGEKLHRA CCGRGRLCREHRVLALPLSSTLPYRDCAPG CILHFPPFVHRYEVDIDEEGKARHTVSLR RIPLTRWKPANPETDPEALLVKEKTMFSGC CNLGDSTANTGSLGNTAKWARVPNYTNM QRLVVPNVGLRCYLLDTRLKGQKKECES PPMIGLRSICMHTKKRVSSFRGNKIGLKDVI TLRRHVETKVRKIRKRVTTKINRHDKIN GKRKTARKQLSLSPCSQCLNLVFLADVW FGFLPSIYLVFLIILYEGLLGGAAVNTFHN IALETSDEHREFAMAATCISDTLGISLSGLL ALPLHDFLCQLS
2692	B	1	678	MKTLLARASRFLALPRTSFNALS KSHNLLG FKDIRSNVEALAQKTQPSVFPKESVQVTPV CYTKGDRESVQKCPLIFRSHSATEQVSIRR GVTVRVAKWRGESHIHGGPDVPGLVLDTS

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				YETSPLIHTPALRVYYIGEDIAMEQVTNLA FPLLYSNSHRVSEPGELGFWGPGESVMPA DAVSVPCTCHPGSYGVQPLVLRIQGYSGT GRWISASAMSCIISDRNG
2693	A	22	334	ALKHFCLCSLIFSVTMTKFLAVLVLLGVSIF LVSAQNPVTTAAPADTVSSLLVLLMMKPLD AETTAAATTATTAAPTTATTAASTTARKDI PVLPKWVGDLNR
2694	A	3	435	RVDPRVRAPRCGDKIKNHMYKCDGSLK DCASDRCCETSCTLSLGSVCNTGLCCHKC KYAAPGVVCRDLGGICDLPEYCDGKKEEC PNDIYIQDGTPCSASVVCIRGNCSDRDMQC QALFGYQVKDGSPACYRKLNRIGNRFGT
2695	A	120	1438	TMNSEDTLRQNLIMGYRQHQAILTAHSTG PRRPAHQSSAEGSLVPCSGNPVPPKG*LW ARQGPAEVSGAGKIPASPKTGFPFLSSH WKLEKGYSPCAQAGCSKGQGLSPQPYLKV LIILGYQA*KGS*FFGSPPSRKVFPSMG TG PQRRKFS*PRFPEGLN*PDCGPGTEPPLGCG CRGLS*VPRSGREKRAMADP*SQLGGSQ GGDFS/*GPEAGRL*VGAQQGPPGVRNRH* SPLLTSS*R/PKARSPDES RGK PQSPLPMS LLP/RGGPSGPHLGPPLEHLPPAPSTPLQNP GPQSMVVGPHSDFYPLPVSPWGSRRLOPTQ LCLPDSKLP GASPPGSAKMAAGQVRWNG NVAR/PTPPGN*PPSSPPGADPLLSQLDPLRP LKWLP SLQFFPKGCGCLGCLCPGPASERSV LSPAPG PGLVGVLGEEQGVARTPGGR
2696	A	2	454	SGHGSSSGTKSSKKKNQNIQYKLGHRRL FEKRKRSLSDYALIFGMFGIVVMVIETELSW GAYDKASLYSLALKCLISLSTIILLGLIIVYH AREIQLFMVDNGADDWRIAMTYERIFFICL EILVCAIHPIPGNYTFTWTARLAFSYAPST
2697	A	506	1317	GRTSSGKAGMWKPGAESWPLHTGAAQV MWFEKLYAGLQCVEKYLIYPVVLNALT VDAHTVVSHPDKYCFYCRALLMTVAGLK LLRSAFCCPPQYLTLAFTVLLFHFDPRL SQGFLLDYFLMSLLCSKLWDL LYKLR FVL TYIAPWQITWGS AFHAFAPFAVPHSAML FVQALLSGLFSTPLNPLLSAVFIMS YARPL KFWERDYNTKRV DHSNTRLVTQLDRNPG ADDNNLNSIFYEHLTRSLQHTLCGDLVLGR WGNYPGDCF
2698	A	86	820	MACYLLVANILLVNLLIAVFNNTFFEVKSIS NQVWKFQRYQLIMTFHERPVLPPLIIFSH MTMIFQHLCCRWRKHESDPDERDYGLKLF ITDDELKKVHDFEEQCIEEYFREKDDRFNS SNDRIRVTSERVENMSMRLEEVNEREHS MKASLQTVDIRLAQLEDLIGRMATALERL TGLERAESNKIRSRTSSDCTDARLHWPVRA ALTSQEREHLSAPKRGLEPWQNILFIQYKP

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				AASSST*
2699	A	3	553	KASVIVHSDVKPFKCKLCGKEFNRMHNL GHMHLHSYSKPFKCLYCPKFTLKGNLTR HMKVKHGVMERGLHSQGLGRGRIALAQ DGVLRSLQEPEFDLSQKRRAKVPVFQSD GESAQGSHCHEEEEDNCYEVEPYSPGLAP QSQQLCTPEDLSTKSEHAPEVLEEACKEEK EDASKGEW
2700	B	123	719	MTEEEWKPMDPSKMRCSFFQNGKESEKE KVPTRSLAQVIPLVNYRGDGSATLQNA DPFVGKAGLGFVDDSPKVEVRCQRLMD NVHKSVCETKKGEAVPALCILLDNPSSC YQPFLEYPRYVKPSSEIPSILPWKENIELGK QATNNSFTEYMLNCAGLDPCHSMCGSRTK IIITCELARNAESQAPPHTY
2701	A	185	284	GQARWLMSVIPALWKAEGGPLEPRSSRP AWAT
2702	A	718	305	SEQEPLLDTPGSREWDILETEEHYKSRWR SIRILYLTMLSSVGFVVMMSIWPYLQKID PTADTSFLGWVIASYSLGQMVASPIGLWS NYRPRKEPLIVSILISVAANCLYAYLHIPAS HNKYMLVARGLLGIG
2703	A	502	822	DSKAAQDLEKLHGVNGMSVDEKPDSPMY VVESTVHCTNILLGLNDQRKKDILCDVTLI VERKEFRAHRAVLAACSEYFWQALVGQT KNDLVVSLEEVQ*FGLCDC
2704	A	313	638	RWRQRWFWCLHCLVLFRTPTFALSQCR PWDDSRSDTSMHSIQWNRMYCNCMQ DEQEADANGKGAQVGDRAWAGR/CR SHRREGTIPGNPHPRAS*RAGWQR
2705	C	431	838	MLLHVGTTAHVAVEHLIGGVQDDEDEM TIGCHGEEMIGDLKNSFGAGGLCIGERVG GPGCCEVLIRMTPTEDVGEERSDMKGIQLS MQERTRCRQFPQGRRHQLGHLLQGGLGRG EAWKYHQIWEEGHWLLREQ
2706	A	244	375	RGMGRTYRGRHTDSRKSDR**GGRRKQTQ KPMSCITVQRKHGTS
2707	A	1606	228	GTSGVQQEISRLTNENLDLKELEKLEKNE RKLKKQLKIYMKKAQDLEAAQALQSER KRHELNRQVTVQRKEKDFQGMLEYHKED EALLIRNLVTDLPQMLSGTVPCLPAYILY MCIRHADYTNDDLKVHSLTSTINGIKKV LKKHNDDEFMTSFWLSNTCARLLHCLKQY SGDEGFMTQNTAKQMEHCLKNFDLTEYR QVLSDSLISQIYQQLIKIAEGLQPMIVSAM LEN*SIQGLSGVKPTGSQKHSSMADEDNS YRLEAIRQMNAFHTVMCDQGLDPEILQV FKQLFYMINAVTLNDLLLRKDVCSWSTGM QLRYNISQLEEWLRGRNLHQSGAVQTMEP LIQAAQLLQLKKKTQEDAEACSLCTSLST QQIVKILNLYTPLNEFEERVTVAFIRTIQAQ



Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LQERNDPQQLLLDAKHMFPVLPFPNPSSLT MDSIHIPACLNLEFLNEV
2708	B	1	468	MQGLVNYQISIKCSNQFKLEVCLLNAENK VVDNQAGTQGGQLKVLGANLWWPYLMHE HPAYLYSWEVRLTAQKSLGPLTSTHSLWG SALCPSPRASGMVIAHTKALDPSQPVTFVT NVTYAADKGPLWEVAAPSSSQRASSGVTE LTRVTPVDLQIE
2709	A	419	2	TSNPKNKVGLLDLELNRLTKALFMALVAH SIVMVTIQGFVGPWYRNLFRFLPLFSYIITIS LRVNLDMGKAVYGWMMMKDENIPGTVV RTSTIPEELGRVVYLLTDKTGPLTQNMIF KRLHLGTVSYGADTMYEIHTK
2710	A	1	570	MSAACGQNYTLALMEMGSVFAGGENKMG QLGLGNLTDITIPSPAIQIYNGQPITKMAFGA EFMMMDCKGNLYSFGCHEYGQLGHNSDG KFIARARRTDGYGRLGHAEQDEMVPPLVK LDFDPGHRVSQIYTGYTCSFAISEVGGLFQ GATNTSRESTTYPKAVQDLCGWIIQSLACG KSSIIVATERAP
2711	A	574	737	AWEGAHVFTTSPSSCHSWVRDYARVGLPP LPLPCPQRALLGLWEVWKGAYSPAI
2712	A	175	2	MALRHLALLAGLLVGVASKSMENITDIDV PAPEVLTRSTAGVRGACASQRGALRCLLG P
2713	B	85	591	MERGPVTCTQAQTVRGRTGHRRRFGPGA HGLREEPEFVTARAGESVVLRCDVHPVTG QPPPYVVEWFKFGVPIPIKFGYYPHVDP EYAEQSCFQAPSFPSPAPBELRVVSARHG LCQALDASWFCTGVQRQPWTQPPTGYHL AQRAGDLYPVGFPGKETYEKVV
2714	A	1196	1459	KQCQRRCLETEVWKLKSLQISTKASNRQD RSTFSAPPRKSQLMW*TSLLSYFQKLPQSP QPSATTALISQQPSTLNPQPWPWGSCPGG
2715	B	1	888	MRIRRWSLMFDVWPMCAFYSWAKASRT FLKADGLPRRKQWVLVEALAGGGVLGVK QITIQLFEVLLRRGKESETYTKMYRRLGP ERCRRSKYAGVERIVDKRKNKKGWYLI RWKGYGSTEDTWEPEHLLHCEEDEFN GLHMSKDKRIKSGKQSSTSKLLRDSRGPSV EKLSHRPSDPGKSGTSHKRKRNPPLAKP KKGYSGKPSSGGDRATKTVSYRTTPSGLQI MPLKKSQNGMENGDA GSEKDERHFGNGS HQPGLDLNDHVGEQDMGECDVNHATLAE NGLGL
2716	A	94	3006	RTRSLTRKAMAEHAPRRCCLGWDFSTQQV KVVAVDAELNVFYEESVHFDRDLPEFGHV LDVHGVHVHKDGLT VTSPVLMWVQALDII LEKMKASGFEFSQVLALSGAGQQHGSYIW KAGAAQALTSLSPDLRLHQQLQDCFSISDC PVWMDSSTTAQCRQLEAAVGGAAQALSCL

475

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				TGSRAYEFNLVCDRKHLKDTTQSVFMAGL LVGTLMFGLCDRIGRKATILAQLLFTLIG LATAFVPSFELYMALRFA\GLLPSLDLASA MSPY*QNGWGPHGGRRPWSWPSATSPSGR WCLRDSPTVSATGGSFSPALRLAYCSSLL LGSARICTLAPDPWEDGRGDTTDPENGLG Q*AETLPGAHEPAGPREDRPLRECPGSVQT PPAPEGDPDYLLCLVCGQSGVLRPEPPSGG LRPGRLSDAAHLWSC*GACPLFQHLHDAE VWPQVEP/RWGPWSWVA*CVSSSSSSQIC PWWSPCWLWWGKWPQLPLPSMCTLPS FSPSSGRQAWGWWASSHGSGASSHL*S CWESTTLSPCSSTAASPSWPA/SLCTLLPE THGQGLKDTLQDLELGPHRSPKSVPSEKE TEAKGRTSSPGVAFVSLGTSDTLFLWLQEP MPALEGHIFCNPVDSQHYMALLCFKNGSL MREKIRNESVSRWSDFSKALQSTEMGNG GNLGFYFDVMEITPEIIGRHRFNTENHKYF KGKGAPGHPMPSLKANFDLLACLRGVGS TLLWPVAVLGAQTRQAGVNEGRSQVADF LRIPVTGCPEQRRNPPSPAPLGTGGPAEER LQFPGVAGSRRGRGRILRAGGIGRASPEG TGAPRPRAGQGRGGPGKPESSGGGPVALR PGDCTCCVLKSQPRQQRGACSAMAFVR LRVRQSVRPRGVIVAALQRPETQGPAPSS ARPDGPESRGLALWRRLRGYASRDRV CNRRCPHAARFPSKRTPSGSPHLHLMSSW AVP
2717	A	1308	369	LRSNHGEDWSQFIGAAQRETTVSLPMPH TWPVSLSTGSCM/TRGTPLPFINNPLQVH FHR/EDDEHSDIAHF*VYFGHWVIMNSHE C/GAWKCEERSNNMPAEDGRVFELHIVLD NEYQAMVNG/QSLLHSFAHRLLPGSVKMV QVWRDVSLSNRCVSSGETVSSSSSFLPPPP PLPLPLLLLLPLPLPDEALFLSLPSHALPSG RCGVLSLCGSHYPQPGGLLQSSAGASGR GAPGVFWQVLVLLTPRGLQGGPPGMRGRV VHKPLLVMELGEQPFSPSVRTATSSASGK APPRCPWPGPRALSPSSVP
2718	A	2	1226	SLGSTISTDWANHYLAKSGHKRLIRDLLQ DVTGVLQAQIIQVVANEKIEDINGCPKNR SQMIENIDACLNFLAAKGINIQGLSAEIKN GNLKAILGLFFSLRYKQQQQPQKQHLSS PLPPAVSQVAGAPSQCQAGTPQQAPGVV TPQAPCQPHQPAPHQQSKAQAEQMQRSLPG PTARVSAAGSEAKTRGGSTTANNRRSQSF NNYDKSKPVTSPPPSSSHEKEPLASSASSH PGMSDNAPASLESGSSSTPTNCSTYSIGPHS GAATKPWRSKSLSVKHSATVSMLSVKPPG PEAPRPTPEAMKPAFNNQKSMLEKLLFN SKGGSKAGEGPSRDTSCERLETLPSEFEESE

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				ELEAASRMLTTVGPASSSPKIALKGIAQRTF SRALTNKKSSILKGNEKGKE
2719	A	103	742	NANTQRRARRREGARLDNLWLEQVISVLPG LVTQGFRCCHSGPMGRGLEPHPIRGAGAGS CQLSIRGRGGRIAPFLTTPRRLAPKGGRDLG FPAPRGTRCLRHSFCRSIARTVT/RTVRGIR GEEARTPGSREMDSVVFEDVDVNFTQEEW ALLDPSQKNLYRDVMQETFRNLASVGKK WKDQKIEDEYKNPRRNLARNYVYHFSLKK WSWSLYARQT
2720	A	1258	586	LLLHSLFPVPRMGNSASNTVSPQEAIPGRK EQTPVAAKHHVNGNRTVEPFPEGTQMAVF GMGCFWGAERKFWVLKGVYSTQVGFAG GYTSNPTYKEVCSEKTGHAEVVRVYQPE HMSFEELKVFWEHNDPTQGMRRQGNHDG TQYRSATYPTSAKQMEAALSSKENYQKVL SEHGFGPITTDIREGQTFYYAEDYHQYYLS KNPNGYCGLGGTGVSCPVGKK
2721	A	2806	382	NEIEKQLNAIRDNIKIGEDRAARLDRKMEE QQVRLNEAEQKYKDIQDKLEKISEETNAR APECMALKADVVAKKRAYNEAEVLYNRS LNEYKALKKDDDEQLCKRIEELKKSTDQSLE PERLERQKKISWLKERVKAFQONQENSVNQ EIEQFQQAIEKDKBEHGKIKREELDVKHAL SYNQGQLKELKDSKTDRLKRFGNVPALL EAIDDAYRQGHFTYKPVGPLGACIHLRDP LALAIESCLKGLLQAYCCHNHADERVLA LMKRFYLPWTSRPIIVSECRNEIYDVRHR AAHYHPDFPTVLTALIDNAVAANSLIDMR GIETVLLIKNNSVARAVMQSQKPPKNCRE AFTADGDQVFAGRYYSSENTRPKFLSRDV DSEISDLENEVENKTAQILNLQQHLSALEK DIKHNEELLKRCQLHYKBLKMKIRKNISEI RELENIIEHQSVDIATLEDEAQENKSKMK MVEEHMEQQKENMEHLKSLKIEAENKYD AIKFKINQLSELADPLKDELNLADSEVDNQ KRGKRHYEEKQKEHLDTLNKKKRELDMK EKELBEKMSQARQICPERIEVEKSASILDKE INRLRQKIQAEHASHGDREEIMRQYQEA TYLDLDSKVRTLKKFIKLLGEIMEHRFKTY QQFRCLTLRCKLYFDNLLSQRAYCGKMN FDHKNETLSISVQPEGNKA AFNDMRALS GGERSFSTVCFILSLWSIAESPFRCLEFDV YMDMVNRRIAMDLILKMADSQRFRQFILL TPQSMSSLPSSKLIRILRMSDPERGQTTLPF RPVTQEEDDDQR
2722	A	1567	1145	AEVLGRAVEPPPGRCWSTPPVAPPARSASA AAMGVQVETISPGDGRTPFKRGQTCVVHY TGMLDGGKKFDSSRDNRNPKFKMLGKQEV IRGWEEGVAQMSVGQRAKLITSPDYAYGA TGHPGIIPPHATLVFDVELLKE

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2723	A	374	656	RRVGCRCFHPSQTGTCT*RPPWNVHH*PAT CHLAYNRHSWSPHRA/HWHIATAIQLSAH VF/ACHYQQLHHYHQHHHHHHHYRHHHH HHHHHYCHHH
2724	A	1171	1639	PMALWADGRARHKVGTCECEGMHPGLKC SGRTLGSQTMLATTPCDSPT*/SNKNGLRS V/SYR*CLINALWLFSPHILVRCGTESS*L LPSLVPSWLP*LVRVR/PLPTGWC*IPSCLKP VPTWSSHSPQRLP*NPATLVCLQNGTARS HSSTPV
2725	A	8	505	GSFKTGLYLPTSDIDL VVFGKWENLPLWTL EEALRKHKVADEDSVKVLDKATVPIIKLTD SFTEVKVDISFNVQNGVRAADLIKDFTKKY PVLPLYLVVLKQFLQRLDLNEVFTGGIGSY SLFLMAVSFLQLHPREDACIPNTNYGVLLI EFFELYGRHFNYLKTG
2726	A	214	32	MTLRMLVPRLLLTRQLVWFFSAATERDPE MMNGIPRKLMSFPSSVTSRRSRRGHHLQS L*
2727	A	2	40	WNSDQPATR*QVGDTGSLPSRKQGHFVLT GIDTYSRSGFAFPVRHAPAKTSIRGLTECRT YCHGMPHCTASV*GTPFTAKKVW*RAHA HGIPRYDHVAHLEAAGLIRWWNGLLKT LQHQLGGDALQGWARVLQEA VYALNQN* V*GW
2728	A	16	444	TPSPSPCPXPRPLAALKPVRLHSFQEHVFKR ASPCELCHQLIVGNSKQGLRCKMCKVSVH LWCSEEISHQQCPGKTSTSFRRNFSSPLL VH EPPVPCATSKESPTGDSGKVDVPVYETLRY GTSALALMNRSSFSSSESPTRS
2729	A	37	655	AEPAAAGAGTLAGDCRAVQGGVHAARPRG AKEGHGPADGHGKGGAGTGQERLAGGAE VCHAQVRGGAAAPGCRVGGVLRAAKAE* GAGRARGRAGIAGGHPAGGHPHQPGQGA G*AEDQGGQRAPGRGEAAGSGR/GA/GPGA GAAGAAAGEGEDQHRPACQAPRRGGGE HEQGGLEVRGGGAGIARGPAGAGRAAG PVAGGAATAGAA
2730	C	257	498	MQKSESGGTQLKNRATGNYDQRTSSSTQ LKHRNAVQGSKSSLSTSSPESARKLHPRPS DKLNPXTINPVHSDDEVFERG
2731	A	342	665	MALDFVNVLLCQLAEVTLGVLREEGASLL VALGSALFPSAAAVGKQGSMTGVTSHMQC PVCQHPRDVLLASPVSHSHACQPQAGCS NCHLGHLTRSPPFQGLLPLLQ*
2732	A	1	825	MKRYSYGSVLFTAFDLGYLDPDEVQQGHE IGRLFDGTEPIVLDLQHYFIDRDGQMFR YILNFLRTSKLLIPDDFKRTLVLPLAAPFS VGLEACPLAGKRLKGSVCPELEFLWKKH RVFSQSLPYKTHAFNEERLQDNKSYHISVL QEPREDTDEGAGAAPDHRSTYKLLSPALS

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LNLGEKNKWLRRYIELLISEREMAAAGSSI PSWTSVSIQVKLRKCQLQLLAKEEVATIVL DETSGVNGIHIEHQLQCLIQVPKLSAPNIAP PTPA
2733	A	135	438	GMGYLHAKGILHKDLKSKNVFYDNGKV ITDFGLFSISGVLQAGRREDKLRIQNGWLC HLAPEIRQLSPDTEEDKLPSKHSDFALG TIWYELHAREWP
2734	A	74	661	HTHKLVAAPRGLPPTSQWPRDAGRQASGG LPSLSTGPPKGPDRDLARGHPAEWLAGSPG NNSPTQGSLLPQLDLYAGALFVHICLGWNF YLSTILTLGITALYTIAGMVPAAGRSTQGT CKGVRRPPPTGPREQPRKWPQEPQKFLP VSLPGARAPSSNLASTGRGPGCCNLHGRP ADAHHGGGGCHPDNQR
2735	A	40	446	RHLLLSLSAVTGKCSFAPDCGELKLPGAAC ACQVVADVSSLL*LCQMRELRCEVATC LGIFGSLGNLLRKEVLHLDWTFKASLLD LICMRSLPGPGTAELLWTAPELLPGPRPG RRTLTGDIFFSTGILQE
2736	A	1	517	LVDPVRGEPGPPSDAVFARDPMRPPGLV RNLQVTDNRNTSITLSWAGPDTQEGDEAQ GYVVELCSSNSLQWLPCHVGTVPVTTYTA KGLRPGEGYFVRVTAVNEGGQSQPSALDT LVQAMPVTVCPKFLVDSSTKDLLTVKVGD TVRVPVSFEHARRPLGPSTCRRTCLGR
2737	A	3	437	NDPRVQKPREEAPAGAAASG*CGR*PGQH PAAA*P*SAGPRRAPTALSPPTAEPSLCPA\ PG*PEQPQCSRRPGGQPRDPVQHRSPAV GPAAGSPLRPCAWSAQRGSPQPDQLPHTP GAAGS*SQLPRPPPSFAQATPSTPP
2738	A	34	576	EELCVREHVTGGICGGSQMMVLLGATTL VLVAVAPWVLSAAAGERRGGESWRRAGG RARSWATGAAMLLGATDAQSGKPSVHFA APKIKPDLGSQINQEKVFWVLSCLRPVAV YGSSGAPGSHPREMAVPELCEFFDSFRETH QILLVYFVCGPRQLFFQCGRPKPKRVDTL ADEACR
2739	A	2	410	CHSTESSDFILPGDYLLGGLCPLHSGCLQV VCSFNEHGYHLFQAMRLAVEINNSTALLP NITLGYQLYDVCSDSANVYATLRVLSLPG QHHELQGDLLHYSPTVLAVIGPDSTNRAA TTAALLSPFLVPMLEQ
2740	A	2	417	STRPEFFGRAPTGLKLLADKNSELFRKYA LFSPSDHRVPRIYVPLKDCPDVVARPKDY ANTLFICRVDWKEDCNFALGQLAKSLGQ AGELEPETEGILTEYGVDFFSDFSSEVLECLP QGLPWTPPEEFSKRRVV
2741	A	1	312	MAPAADREGYWGPTTSTLDWCEENYSVT WYIAEF\SWLMGFLPTPSSLRDLTASRWV RSLPPSRSPAGRQPGPAEELPKASPCPWGK

479

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				SLSRPFASFSAASSGPS
2742	A	2	374	FRDLQCALYNGRPVLGTQKTYQWVPFHG APNQCDLNCLAEGHAFYHSFGRVLDGTAC SPGAQGVCVAGRCLSAGCDGLLGSGALED RCGRCGGANDSCLFVQRVFRDAGAFAGY WNVTLIPEGA
2743	B	218	656	MGPVPLVWAMSQSLSAKMDRRRTGVM MTSTPITWGTLEKTMQEAELLERQGQTK TPDSMFLAMEESLNVTFVKNITTQFMVCG FNPYVFLAAKADQLQVVVSHTTTASQER
2744	A	85	396	MILINFREICLKVLTHTPLCVSGGCVLLYILA LTCCYTNSLLISHLPPLSLPTETQTHLFMYR VLKVRKDKNHVFHTYLVAKETETYGEE LIPLPCREHQD*
2745	A	1	3899	NRPSSASSTSSKAPPSSRRNVGMGTTRRLG SSTLGSKSSAAKEGAGAVDEEDFIKAFDDV PVVQIYSSRDLEESINKIREILSDDKHDWEQ RVNALKKIRSLLAGAAEYDNFFQHLRL DGAFLSAKDLRSQVVEA/CITLGHLLSV LGKFDHGAEAIMPTIFNLIPNSAKIMATS GVVAVRLIIRHTHIPRLIPVITSNCTSKAVA VRRRCFEFLDLLQEWQTHSLERHISVLAE TIKKGIHDADSEARIEARKCYWGFHSHFSR EAEHL YHTLESSYQKALQSHLKNSDSIVSL PQSDRSSSSSQESLNRPLSAKRSPTGSTTSR ASTVSTKSVSTTGSLQRSRSDIDVNAAASA KSKVSSSSGTTTFSSAAALPPGSYASLDGTT TKAEGRIRTRRQSSGSATNVASTPDNRGRS RAKVVSQSQRSRANPAGAGSRSSSPGKLL GSGYGGLTGGSSRGPPVTPSSEKRSKIPRSQ GCSRETSPNRIGLARSSRIPRPSMSQGCSD TSRESSRDTPARGFPPLDRFGLGQPGRI SVNAMRVLSTSTDLEAAVADALKKPVRRR YEPYGMYSDDDANS DASSVCERSYGSRN GGIPHYLRQTEDVAEVLNHCASSNWSEK BGLLGLQNLKLSQRTLSRVELKRLCEIFTR MFADPHSKRVFSMFLETLDVDFIIHKDDLQ DWLFVLLTQALLKKNGEADLLGSVQAKVQ KALDVTRDSFFPDQQFNILMRFIVDQTQTP NLKVKVAILKYIESLARQMDPTDFVNSSET RLAVSRITWTTEPKSSDVRKAAQIVLISLF ELNTPFTMLLGALPKTFQDGATKLLHNH LKNSSNTSVGSPSNTIGRTPSRHTSSRTSPL TSPTNCSHGGLSPSRLWGWSADGLAKHPP PFSQPNISIPTAPSHKALRRSYSPSMLDYDTE NLNSEBIYSSLRGVTEAIEKFSFRSQEDLNE PIKRDGKKECDIVSRDGAASPATEGRGGS EVEGGRTALDNKTSLLNTQPPRAFPGRAR DYNPYPYSDAINTYDKTALKEAVFDDDDME QLRDVPIDHSDLVADLLKELSNHNERVEER KGALLELLKITREDSLGVWEEHFKTILLLL

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LETLGDKDHSIRALALRVLEILRNQPARF KNYAELTIMKTLEAHKDSHKEVVRAAEEA ASTLASSIHPEQCIKVLCPHQTADYPINLAA IKMQTKVVERIAKESLLQLLVDIIPGLLQGY DNTESSVRKASVFCLVAIYSVIGEDLKPHL AQLTGSKMKLLNLYIKRAQTTSNSSSSSD VSTHS
2746	A	153	1224	RVFSESVCSFVRNLEFLWRFAFPLAPAGRC PPGVPLQTSRDTDAHRSSPLPPARASPGQ VAAAYRWARCPGCGGRKPRSSGSWQLCR CPTLPPPPRGRSSSGRC/RTWSPSPSCFPHFQ SGPRTRAPTPSTTPGYSGSYSSGPGR*GLS PLHAA/VSPPLPPGGP*GSWARAGLGSIAA HSPCPLCRSLIRSR*QTCTRSPT*NCEVPPS AP*AASPLRTMFALVRTAGLKVHLLPLGY CTTMS*SSSMPQTVPVVVKVSNIPSVHPP*P CCKDCTISRSRSIFTRSPICNPPGFLLPFCSPS TGQ*SL*KEPPLASWTHFRSDVLLLFVSVM NGSTLSLGCPSQKAVIALVQVT
2747	A	1	996	MKIHSACFVIEQEEKKKTEAHKEGDGVKR ADKILGVTKDPGTIAGLNVVRINEPTAASI AYGTDKKFGAERHVLIDLRDEIFDVSVLT LEDEIFEIKSTAGDTHLGEEDFDNQMINHFI AEFKYKHKDSRADIYTSITHAQFEELNAVL FRGTQDPIELALQDTKLDKLIHVIVLTQTF TTYPDNQPDVLIQVYEGESAITKDNLLVI QGFELTGILPAPFAVPQIKVTCDIDVNSSL NISAVGKSTEKENKIITNDQGHLSKEDIEN MVQEAHEYKADEKQKNKVASKNSLDSYA FNMKATEKLQGKINNKKDKQKILDKCNKIIN
2748	A	73	1210	IPPPSSPSPAAAPRAQLGKDALSPLALLR PRRAYPRPLPTSESLAWGSPPSRFGSPAS QPRSPRLSFLVLGVACSAILMYIFCTDCWLI AVLYFTWL VFDWNTPKKGRRSQWVRN WAVWRYFRDYFPIQLVKTHNLLTRNYIF GYHPHGIMGLGAFCNFSTEATEVSKKFPPI RPYLATLAGNFRMPVLR EYLMSSGICPVS RDTIDYLLSKNGSGNAIIVVGGAESLSSM PGKNAVTLRNRKGFVKLALRHGADLVPIY SFGENEVYKQVIFEEGSWGRWVQKKFKQ YIGFAPCIFHGRGLFSSDTWGLVPYSKPITT VVGEPITIPKLEHPTQQDIDL YHTMYMEAL VKLFDKHKTKFGLPETEVLEV N
2749	A	351	205	DLYSEKASADHEGAEQFTDEFKVIADGN LMPEQVYNVKTSLFWCMVP
2750	A	172	2	MLEQASLWLGSRFLLAGFLVSSSCPSLEQA AKGEGCSPICFAHCLDSLVRNFLCHP
2751	A	2	1410	GPLIDLCKGPHETHTGKIQTIFTNSSTYW EGNPEMBTLQRIYGISFPDNKMMRDWEKF QEEAKNRDHRKIGKEQELFFFHDLSPGSCF FLPRGAFIYNTLTDFIREYHKRDFTEVLSP

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				NMYNSKLWEASGHWQHYSENMTTFEIEK DTFALKPMNCPGHCLMFAHRPRSWREMPI RFADFGVLHRNELSGTSLGLTRVRRFQQD DAHIFCTVEQIEEIKGCLQFLQSVYSTGF SFQLNLSTRPENFLGEIEMWNEAEKQLQNS LMDFGEPWKMNPGDGAFYGPKEIDIKIKDAI GRYHQCATIQLDQFLPIRNLTYVSKDGDD KKRPVHRAILGSVERMAILSENYGGKWP FWLSPRQVMVIPVGPTCEKYALQVSSEFFE EGFMADVLDHSCITLKKIRNAQLAQYNF ILVVGEEKIDNAVNVTRDNKIHGEILVT SAIDKLKLNLRKTRTLNAEEAF
2752	A	319	495	MVASFRESRVLLGLVVRVLTDFLTQVV RVGSECGDELVRLYSFTDEKANYLQQGGC R
2753	A	23	1255	LRSIYTHYRESVPA/HLTDSFPDLLGLAA ED*HCPIALEAL*TTDAELRVTLTVEGKPV PFLINTEATHSTLPSFQGPVSLASITVVIGDG QA\SKPLKTPQLWCQH*TIRRFKHSFLVIP\ TCQVPLLGVEDTLTKLSASLTIPGLQLYLIAT LLPNPKPPLCPPLV/SPQLNPQV*DISTPSLT TDS
2754	A	277	467	GLGPHDYLYSILSIERSCCC*CCCCCRRRR CCCC/CV*GCSRFLCSIAESTPSGALRRLR GGR
2755	A	86	593	ASALLFVVGFAESLREFTADCPYKCPVAP EPLPQPLS/PLQCPGEESTDSPFLPTVQPVK SRCSPFIEESPRANRSIPAFGSHLECASSSR SFHGPPPCCLWGLPLSAPSPHVLHPPASAAI GPACCVTSLCPGAPQAQRPRKVDQTSSAP GAGPGTQDGNERPNP
2756	A	3	3617	YWKERPTQKVPRATENHGLKSYLQKTKL SIDEAAFLLPDTNLKSELLELLTHWLQVGV PMTPSLGSINLLGWLTELREHTYICWFTV KETTRDTDEEMCRTEPALACSISHYCDDGC IQMLNTPETLQCSAKDSKHFPKECSIPGEN RPPSDTGKTVKFLSLNIFNLQLAESTDAEQ RANCILRCFLTETTLNYQKILSVRPGTKLAT ASHVSGGLQTPPFGLAQHLIRPHAFLAPK DPLTSFTEENSRSGKTRCRSKKCAMRVVK SYSAILPKKRESVLTKTLLVAPTNEQTDPV LRMCCGKTGLKKGAGFTLESRGQRRMRA GCPTLCVRARVTETDPSICSEVTFWSMILM LMDVCQCLGIEFGIYCSLRSLDLFVPIFLE KVFQVFEGTSSPIMLWFLQTHRGTTLVALD KIQKNSLDYQAETLVLPYFLPNKWNLSVF AEPPGTGDVVMQAPLWPPPLGLYWALEH YDQHVAKPARQRSLSLWPPPTAHKGFLQ GHCQCSLKTQRLFSQLMANAARPETQASG QWTFSPGQIQKCSPRSRLALGTTPRACLLL YPTVAELGSTEFNVKPSICCTLPYQGAQSPS



Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LHTLQLRGNGVGGQHQFQKTVSLDPFNAS FRDMKLLKLGKSGISSWFVSIAAAVGDEGL VPRSMELYSQKAYDCLCCVMQVVRKVGE SWQSQTSPSSHTTQKANLTSTLPPTTALS FPGSGYQEWGTAVKILESMEATLEQDNKT RLEQFGGFRKEDRKMWESLELPRDLWN DFDQNA DSDMDNEVQAEVVS DGDKELVR NWSKVWKG NVGLEPRYRVPTGALT SRVV RRGPPSFRPQKCRSTDSLHHEPGKAAGTQC QPVKDLPKAVGAHSLHQPALDFRQEYLN FSKNAKFQYECGNYSGAAENFYFFKGLVP ATDRNALSSLWGKLASEILMQNWDAAAME DLTRLKETIDNNDKPSFTHVVGKERYLN AIQTMCPQFFRY/L*LTAVHNKQGIVRKRR PRV*KI*LSFIKQESYTYKRPNLQNLLECL YVNFDFDGGSRKS*GECEPGLVNDFFLGG *S*GFQ*KMPRLFIFETF/CRIPPSVSAINML AD\KLNMTPEEAERVDW*NLIRKWQAWM PQDLIPKLGSCGLWGNNAV\SPLQQVIEKT KSLSFRSPDVGP*IMRKNLNQNSRSEAP*R GQLQDSGLLLKNHKEKMKKKNYQRKMK
2757	A	1	3090	MHKELPALAACGLVADFPVGEETADFG PLVLDSDSDSDVDRIEBAIQEYLKVGSSK DQGSASPVMSRADSFEQSIRABIEQFLNEK RQHETQKCDGSVEKKPDTHENSAKSLSKS HQEPATKVVHRQGLMGVQKEFAFCRPPR LAKTNVQPRSLRSKVTTTTTQEKEGSTKPA TP/TRPSEAVQNKSGIKRSASTARRGKRVT AVQAPEASDSSDDGIEBAIQLYQVQKTHK EADGDPPQRVQLQEERAPAPPAHSTSSATK SALPETHRKTPSKKKPVPTKTTPGPGDLD ADHSPKIPKETKAPPPTSPASRSKFVWSSC QADTSAELIAVLDFKTL/PAPMEGSDGSL SASPLFYSPNVPSRSDGDSSSVSDSDSIEQEI WTFLALKGTASEAPGGEGAARVPGDTRTS QGQGKTDEARHLDKKKSSDKSSSLDSK DLDTA IKDLL/RRVPGPSSQPWLLV*QQQFS GQRR*HRTGD*EVFGGKGQGVGSPRPGPA LSLEAHTCWRRRTAITGQAGRC/LCYDSQD PKCGDLKKPSKKRVKRKPYSTTKVTSGSTF NENTRRYAVHTNQCRPHGSRVKKKRYP QEDDFHHTVFSNLERLDKLOPTLEASEESL VHKDRGDGERPVNVRVVQVAPLRLESSKY TGITCQENNLDAKKAPHEDTVHDITNEDA THDIANEDTVHDIANEAADKGIANEDAAH GIASEDAAHGIASEDA AQGIASEDA AQGIA SEDA AQGIAKEDAAQGIANEDAAQGIANE GAAQGIKEDAAQGIKEGA AHGIANEDA AQGIANEDSAHGIASEDA AHGIAIANEDAI YDIANDTVQGTLTRTLYTTSLMRTPYKAL VMRTLYMTSLTRTLYKPSLTRTLYTTSLM

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				TAPYKTSPMRALYTTLMIPTRHANADTV HDIANEDSVYDIANEGAVYDIANDTVQGT LRTLYTTSMLRTPYKASVMRTLYTTSLTR TPYKPSLTRTLYTTSMLTAPYKTSPMRALY TTLMIPTRHANVDAVHDIANEDTV
2758	A	1	1026	MTLGPLTNQRKEHLTNFKSVSTPSSSEFEC FFSTDSSDLSPSPQAARRQAEPGACFKCWK SGHWAEECLQPRIPKLPICVGPVHWSDC PAHLAATPRAAGTLAQGSLTPSQIFLAEWL KTD TARSPQKPPGPSQTLWVTLTVEVAAT ALILLEALKITSYAPLTLYSSHNFQNLFS HLTHLSAPKTLQLYSLFVESSTITTVAGPDF NPASHIIPDTPDPHDCISLIHLTFIPPHISFF PVPHPDHTWFDGSSSTRPNRHTPAKAGYAI VSSTFIEATALPPSTTSQAKLIALTQALTL AKGLLVNIYTDISKYAFHIQYHHA VTIWAER NFLTT
2759	A	1	383	TRKCGQLPRSVSLSPGQPLPGSVRHRPV LRRPLPRAQGSSSSFRPRPFAPDTMDKFW WHAAWGLCLVPLSLAQIGECPPQPGQDQ CGVLSADPAAAPPAESALGDWSQVSCLS ALGSGKQGW
2760	A	1057	1226	ARPSRVEAQMLGARRAASWLWAPWFCPN EG*NQPGQHSETPSLQKVLKPGMVV/HLL WSQLGSLRWEDRLSPGD
2761	A	349	1	NQTPFFFFFGGTETTTSTLCSYGLLILLY PEVA/ESASQRDPWEAAVWRWLEGPGSA QPPSAPAKGQELDPVVGQRPVPSDDHVQ WPYTNAVLLIEIQRFISVVKRTLTLDTLY
2762	C	199	531	MTGIVAKQNSASVPLPARLVRTVNRKLL GAGTGSPLRKEARRERFLDGDQDGDGPR QPSMGLPHKQVQNRAMAKVVITFAPTNA MQLARSPKTLNFMKIIGEMESVLE
2763	A	1	1428	MVNPTVFFDTEPLGRISFELFADKFPKTAG NFHALSTGEKGFGYKGSFHRIVPGFMCQ GGDFTCHDGTGGKSIYREKFDKDNFIRKHT VSGILSMANAGPNANSSQFFICAakteWLD GKHVVFSSKVKEGMNIVETMECFGSRNGKT KGAGLAGSHSQRWLAASVCGASQPSRLLS TACRQQLQISGRSKGCSRKTSGLDQGLT KDGTNNTQGIKLQLGEEEEHSRPSLVPV SQLKANGSSSASICAEDGPAPVPGCQCQ NQGHQNKRPRTSQLCQMPKTHLVVADA RPNISRVFFGLPERESALWSFPRDWLVNLL NQCDLGIRNQFEVEVLSYGHLPAYSARC FTARSEDPRKDECETCCIKYPNGRNVLSQE NQQVFVLNGIQTMSGYVYNLGNELASMQ GLVDVVRSLSPQGTDTFAMLD AFRANENG AAPLPLTANSDCNGYWRRLADFECTWAH SQGGCHA
2764	B	159	2657	MTCGTDGAITFWESLTGHRYIHKPTNPDEP

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				PVAEQPKPLYPYRTIGCVFNHQMFLGNCQ PSDAVETCVFDLNDSEKWKPMSEEAIKSV CAPGATTSLPPFPLCASTIDASVTSNEIEM QLRLLVSEHRKYTKIHTCPSPTGGPVEPAD TKSQPSVCMDFTSHEYRISDPFLVEKNLPK EKTANTAGHQKEQTGDTLPLRNITGTVRV HGFILEVSETKNPPNPGHKTSISQRPKALV SLGPEVRRGTRGEDEKALEKEGGRRWEC GGANELCGRPPAFTRVTVHWGKNDQTF QDLLDTGSELTLPDGPKRHCAPPVKIGAY GGQINGVLAQVQLTVDAVGPWTHPVVFP SARMHNWNRHTQQLAESHIGSLTVHLSSD PKGCHSEWGPEQEKALQEVQAAVQAALIL EPYDPAGPVVLEVSLADRDVWVSLWQAPI GESQQRPLGFWSKALPSSAAIKRVMHSSIP SSNGSGIYMIGLEQVRKAQIVLHDMQPPCE NGTASALQPLSRKSLKDSSEGKSSQWAE RAVHLAVHVAWKEKWPDVRLDTSWAV ANGLARWSGTWKEHDRKIGDKEVWGRGT RIELSEWSKTVTIFVSHCFYQDYHPSVGSQ NALYTNMVFHTALPLTKALTLRLKNCNSG LMLTEFTGLTMFPIIQGWGKVLQKAVYAL NQRPIYEWKEESCLHTGVADALRGNWAE GHREHKALWLGLWSTWSQHPLRLSKTTR HHPGLGLVSEDICEAGGATEELSRASGFAT GYGKRKEDTKKHKQHSVSDIM
2765	A	3	662	TRIAETILKKKTKVGGTILSDFKMNKARVL EIVWYLWSNRCMNQWNRIEDPETDPQTN GALAIGHPTKQIKLTNRPQSLNLRPDM KMNSKWIVDLNVKCEAIKTF/EKKTRENH HQBHNLEDNTYKLNFKJCSAKSAV/SRIKK K/PTA*EKIFANRLSNIGLISREYKQLKLSS *KTV*LENGGLAWWLTTPVPSLREAKVDEP LEARGSRPAYPTW
2766	A	736	927	SVAHSSCVSHTMHMTLLGRRATINCLFRN GRGQVQWLTSAPALRKADVGG*LEPRSS RPAWAT
2767	A	194	3	MVMLTLAIRLMQFEFRQFFIKVNFRMRGL SKMAMLLLCRAPYSYKKEEGWSVLSGY FLTAGNF
2768	A	593	230	DFYLYPERKKRGQMMAVSLTTRPQESVA FEDVAVYFTTKEWAIMGPAERALYRDVM LENYGGCGPL*CHPTSKPALVFSLEQKES CFSPATGSSLSRNDWRAGWIGYLELRRYT YLS
2769	A	3	4804	KRLNIQKTLVAFSEAVVMQPSVLLDD LDLIAGLPAVPEHEHSPDAVQSQRLAHALN DMIKEFISMGSVALIATSQSQSLHPLLVS AQGVHIFQCVQHIPPNEQRCCEILCNVIK NKLD CDINKFTDLDLQHVAKETGGFVARD FTVLVDRAIHSRLSRQSISTREKLVLTTLDF

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				QKALRGFLPASLRSVNLHKPRDLGWDKIG GLHEVRQILMDTIQLPAKYPELFANLPIRQ RTGILLYGPPGTGKTLLAGVIAESRMNFIS VKGPELLSKYIGASEQAVRDIFIRAQAACP CILFFDEFESIAPRRGHDNTGVTDRVYNQL LTQLDGVEGLQGVYVLAATSRPDLIDPALL RPGRLDKCVYCPPDQDGSSSSDLSLSS MVFLNHSSGSDSAGDGECGLDQSLVSLE MSEILPDESKFNMRYLYFGSSYESELGNGT SSDLEDESMNQPGPIKTRLAISQSHLMTAL GHTRPSISEDDWKNFAELYESFQNPKRKN QSGTMFRPGQKFFDEITELTYLPSFHHKAA PHQAEPGNSSASAPPPYNPFITSSPHTQS GLQFRSVTSPPPSAQQFPLKEVAGAKGIVK TALETAPTLALPVSSQPFSLHTAEVQGCAV GILTQGPGPCVAFLSKQLDLTVLGSPSCL HAVASAAALILLEALKITNYAQLTYSSHNF QNLFSFSLTHLSAPRLQLYSLFVESPTIT ILPGPDFNLASHILDTPDPDDCMSLIYLT TFPFHISFFSVPHVDHIWFTDGSSTRPDRHS PAKAGYAESSTSIIEATALPPSTTSQQAELI ALTRAFTLAKGLHVNTYTDKYAFHILHHH AVIWAERGFLTQGGSSINASLIKTLKAAAL LPKEAGVTHCKGHQKASDPITLGNAVADK DRTIDGSSQVIEKNHNGYSVIDTGTLVEA ELEKLPNNWSPQTCELFALSQALKYLQNNQ KTISILIQKEPSALGLTPERKGNVGHAGKG PLESSSPDPFLCGQERREKGCRTATSVSITN PINRGPVVVTHTPGKELTPEHKGNVGHAGR DILAKAGAIHLNIGEGTPVCCPLLEEGINPE VWATEGQYGRAKNARPVQVKLKDSTSF YQRQYPLRPKAQQGLQKIVKDLKAQGLV KPCSNPCSTPILGVQKPNRQWR/TLCHQAT QALFNFLATCGYMSKPKAQLCSQQ/RYL GLKLSKGTRALSEEHIQPILAYHPKTLKQL RGFLGVIGFCRKWIPRYGEIARSLNTLIKET QKANTHLVRWTTEVEVAFQALTQAPVLSL PTGQDFSSYVTEKTGIALGVLTQIRGMSLQ PVAYLTKEIDVVAKVVAVALVSEAVKIIQ GRDLTVWTS HDVNGILTAKGDLWLS DNC LLKCCALLLEGPVLR/LCTCATLNPATFLPD NEEKIKHNCQQVISQTYATRGDLLEVPLTD PDLNLYTDGSSFVEKGLRKVGAVVSDNG ILESNPLTPGTSAQLAELIALTWALELGEEK RANYTDSKYAYLVLHAHAIIWKEREFLT SERTPIKHQEAIRKLLLA VQKPKEVAVLHC RGHQKGKEREIEENCQADIEAKRAARQDP PLEMLIKQLV
2770	A	1	2919	MLLATALRGFLKNGDRGHVDTEEWSYP WAASFGQLRSSQNCPGASASGRTGVPTVL VARTDADASDLITSDCDPYDSEFMTGERTS

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				EGFFRTHAGIEQAISRGLAYAPYADLVWCE TSTPDLELARRFAQAIHAKYPGKLLAYNCS PSFNWQKNLDDKTIASFQQQLSDMGYKFQ FITLAGIHSWFMFMDLANAYAQQEGGMK HYVEKVQQPEFAAAKDG YTFVSHQQEVG TGYFDKVTTHIQQGGTPDKAFTPHPAKPAH KPGEQPMKNNPLISIMPTWNRQQLAIRAI KSVLRQDYSNWEMIVDDCSTSWEQLOQY VTALNDPRITYIHNDINSGACAVRNQAIML AQGEYITGIDDDDEWTPNRLSVFLAHKQQ LVTHAFLYANDYVCQGEVYSQPASLPLYP KSPYSRRLFYKRNIIGNQVFTWAWRFKECL FDTELKAAQDYDIFLRMVVEYGEPPWKVEE ATQILAINHGEMPIHSSREHFRVLPFCRSTR PFRQARKISRIVVTSTKSDSLYTVGMLALS VRAIRCPLYLLTGLISVSKNGLWYCELQVA LHGRSVTLYEKAPLSEQCSKKAHDQFLA DLASILPSNTTPLIVSDAGFKVPWYKSVEK LGWYWLSRRMQIBETFRDLKSPAYGLGLR HSRTSSSERFDIMLLIALMLQLTCWLAGVH AQKQGLDLGVYGA PETFLIDGNGIIRYRHA GDLNPRVWEEEEIKPLWEKYTLATIDVLQF KDEAQEQQFRQLTEELRCPKCQNNSIADSN SMIATDLRQKVYELMQEGKSKKEIVDYMV ARYGNFVTYDPPLTPLTVLLWVLPVVAIGI GGWVIYARSRRRVVRVPEAFPEQSVPEGK RAGYVVYLPGIVVALIVAGVSYYQTGN YQ QVKIWQQATAQAPALLDRALDPKADPLNE EEMSRLALGMRTQLQKNPGDIEGWIMLGR VGMALGNASIA TDCYATGYRLDRTTVML DGDR
2771	B	1	1773	MALGISAPVALQGTAPLLAVLSGCSFPKH MLQTVNGSPFWGLENGGPLLRLRAGSAPV ETLELFSSLNKILHSYHSSVVKCDLILLGRW TKAWDPLSAGGGCHTGPLPLQVEGNHPTG SYRVPNRPQYRSVAWGLGTSGLVNYTFLL NSGETTYQFLRGNKDFLKNHIKLN YCFLLI EVDNLT LVFVIEKTLGQIFDIPKVELLFSYQ CFPMVENRQKPEGEEDCVIQLSELSCTECS KKAWRMEVLHTNKT TNATQCGGPAQLQQ FNAVLSEKVHIVPSLLRSWNIISHGRFSPFE TFNTKNCIAYNPNGNALDESCEDKNRYIW LEKPOETYSNDRRESKHIPLRMAAERRRAE QKEKYPLIKSSDLGASEAIRQRQSSAAKLR KSGKESVREPWARVPGALGVAARALIAED AGLSRVILFHYGESWNLLRADQRLIFAKS WPRASRYQQGHQDLFILRSDLPSQVFIRDK LMERRNRR TGRTEKARIWEVTDRTVRTWI GEAVAAAAADGVTF SVPTPHTRHSHYAM HMLYAGIPLKVLQSLMGHKSISSTEYVTKV FALDVAARHRVQFAMPESDAVAMLKQLS

487

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2772	C	148	306	MRPCCWWATLCGKHLRMC SHALKMRPN ASAAETEQLNAHSRGLMNSSSRPAP*
2773	A	2874	3062	GNRAGALPGATLLILAGFLPSAHQNRPSRN PVSRRPNTQRRARRKHYALADGYTERRWT NAP
2774	A	1	660	MPNFFIDRPIFAWVIAIIMLAGGLAILKL PV AQYPTIAPPAVTISASYPGADAKTVQD TVT QVIEQNMNGIDNLMYSSNSDSTGT VQIT LTFESG\QVQNKLLQAMP/LLPQEVQQQGV SVEKSSSSFLMVVGVINTDGTMTQEDISDY VAANMKDAISRTSGVGDVQLFGSQYAMRI WMNPNELNKVERNSRRQDVGERDISSGSR KVNKESREDEEVT
2775	A	78	264	PVERSNLGVRLYACCGLLLRPAYPQHFAH GYVDKIPDYPRRAGTLTGLHPMQVCRCRR AREL
2776	B	1	921	MLDDYGGSLSELAREQLPAAEQAALAQLA ARSLAPVPDDTGAGMSNDTPFDALWQR MLARGWTPVSESRLDDWLTAQPDGVVLL SSDPKRTPEVSDNPVMIGELLREFPDYTWQ VAIADLEQSGRIGDRFGVFRFPATLVFTGG NYRGVLNGIHLAELINLMRWLVPEPQQL HQPLTTVQNANDCCCDGACSSPTLSENV SGTRYSWKVSMDCAACARKVENAVRQL AGVNQVQVLFATEKLVVDADNDIRAQVES ALQKAGYSLRDEQAEEPPQASRLKENLPLI TLIDSSYFPHGTALAF
2777	A	47	275	FPCPPAPHVCGPPPCPRAFPVGQSSSQPV ATGFP*SPVCPPPRLYWGPGTERHWVETH YRAFLPSQHLSSPVTAA
2778	A	749	1020	VLVRDPSQPAQPFVSFSPQKHRDEKLYFL PKGVSGGSELGRQPYPYLPVSPITLCPWG HLSLAPPSVPPTACESSSELWPSLSWTWAE
2779	A	271	86	MPLHTCLVHVGVSHAARGSPVCPVVLWV WFCVHFQVIHMWAHECVQADVWAHIQD CAQVCV*
2780	A	3	523	AAANRKR AAYSAAGPRPGADRHGRYQL EDESAHLDEMPLMMSEEGFENEESDYHTL PRARIMQRKRGLEWFCVCDGWKFLCTSCCG WLNICRRKKELKARTVWLGCEKCEEKH PRNSIKNQKYNVFTFIPGVLYEQKFFLNL YFAVISCSQFVTALKIGYLYTYWAPLGF
2781	A	2	141	EQFLRRQIASEKEETERLKAIEAIQSRQQH GRSETEEYSSLLQF
2782	A	3	402	GNGGFVVHVLNNKEFHFTSSTEVMHQLR KLSDKQVDHENDDADREDEEHSQEDRER GLHMKLDHDLSDRESEAGTSSEHEDGE REGSPRTYSRLSVPMLPTVLLDRKIETLLT EWNKNPDMLFTIHPMY
2783	A	333	695	ISVFRSPGQSTSQHDAATWPFLHISGEGTP SRRKAPPAFHPHTQACPSTCYCHTLASRRG

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				PCNGRYHRPVYHPHTAMQRDPPAGPRGCQ SPCWHYTPACRHPGCRHYR*HGQHDPWP Q*HC*FGSPGQSTSQHDAVTWPFLHIPGER PTASRRKAPPAFHPHTQACPSACYCHTLAS RRGPSNGRYHRPGYPHTAVQRDPPAGPR GCQSPC*HEPPACRHPGCRHYR*HGQHDP PWQ
2784	A	91	297	MSLVKLFNLLVFSYRRGAVITIKIEVKIKVT YVKCQAHGERLINGHYDYSACHVIKLMFC ABEKKPHQ*
2785	A	2	103	TGEKVVPGEVNPPNGPVGDPDLSLLFGDVT LKSFDLSLTGCGDIAEQDMSMTDSMASG GQRANRDGTRSSCLVTYQGGGEEMALP DDDDDEEEEEEEVELEEEEEVEKEEEDDD LEYL*EGSTRRGKPTQWPCGGPTEPLVWG CDIPEKL
2786	A	24	332	QPQYIAPLMANFDPVSRSNSTVRYFDNGT ALVVQWDHVLQDNYNLGSFTFQATLLM DGRIFGYKEIPVLVTQISSTNHPVKVGLSD AFVVVHRIQQIPST
2787	A	210	281	FHHKQLHNPVLECHQPAGPCHYL
2788	A	2	1211	WTPPGAPGAKGPRQGGCCSGLLRPPRVSG KTCGARPPWPWRSLSRIPKREGLGEEDTA VAGHELLLPNERSFQNAAKSNNLDLMEKL FEKKVNINVVNNMNRTALHFAVGRNHL AVDFLLKHKARVDVADKTRMRELLLEIFL TVPRAQFHDHLHCLESKLEDCEMRDTRLHM QAVYRETNILTHTVTCVRLGALSYLKTMA CRPQQNILSDKNMDSVLTSMYMLGKLNHL SVLQFLYLKNEKDNSTYVNLILSERIPTLIF QIQPKPYREVMQLAQLVVLALTLSFTV VVLNSIRAMVPSEIFKAKDLLSRKIHIIY DKNIAYESAVPIMPVIPQTGSPTYTSSAALP QCLTPGNTIHSVAIVNGSSWSSALRSQCDH RLHTCSFTLVQRPHTQLI
2789	A	1	334	FRANRTVKDAHSIHGTNPQYLVEKIIRTRY ESKYWKEECFGLTAELVVDKAMELRFVG GVYGGNIKPTPFLCLTLKMLQIQPEKDIIVE FIKNEDFK*VQCCLANIRGMY
2790	A	3	1794	AMLPMELGCGPLPEPLPVGCSRFSLFK*QT CISTVP/GYMVTAQSMSTPPPPSPSTLPSSP SPPPPLPQPLPPPPSPPTLSSLSPPPPPL VSPSTLPSPQSSPQLLPSSSPSLPSPPP SPPLPSPSPAIPSLPPSPQPLPPPPSSPPS LPSPLLPPPLSSSPSSPLSPSPPPSPPSLPP SPPPSPPPPPPPQPPSPSSPLSSPLSSSQPSL LPSSLSSLPSSSPPLPLSLPLSISPP*LSLL SPLPPSPSLPPSSSQST*TIGQCFSL/VMWHV APCTYLALAGNTLMWPLMSASSKASGG VSMFVWRNVEPCSVAVFSWYSVPFLTPC SRVRPSNLPVTQWPPTRAKNLPQRLLTS

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				VHQAQSLSALCKEQDSSSEKDGRSPNKWD KDHIWWPMSGGHDLLQQAAPGPGRHQGH PYQDNWTISQILSERWYTLGPNEMQKYHD LAFQHMAGEDIASDEEHMVIHEEGVMVS LLMTALAPLTLISSSRIFGKVYGPSPSSYT YDASSSTLAPTSFLLGPGAFKAQESGEEA EDGLRELETEKALSSSLRRALDQ/*LALIM QLFQAHCFLLST
2791	A	230	2579	AICDPCYWRMEKSPRMMEKKLSKGMIPD WESRWENKELSTKKDNYDEDSPTVIEK VVKQSYEFNSKKNLEYIEKLEGKHGSQV DHFRPAILTSRESPTADSVYKYNIFRSTFHS KSTLSEPKISAEGNSHKYDILKKNLPKKS VIKNEKVNGGKKLLNSNKSAAFSQGKSL TLPQTCNREKIYTCSECGKAFGKQSLNRH WRIHTGEKPYECRECGKTFSHGSSLTRHLI SHSGEKPYKCECGKAFSHVSSLTNHQSTH TGEKPYECMNCCKSFSRVSHLIEHLRIHTQ EKLYECRICGKAFIHRSSLIHHQKIHTGEKP YECRECGKAFCCSSHLTRHQRIHTMEKQY ECNKCLKVFSSLSFLVQHQSHTTEKPFECQ KCRKSFNQLESLNMHLRNHRLKCDFYLM NAIYVGKPLVIGHPCFNTEFILERNLTNVL NVGRPSAVVQTLPIREFILEKSHINVSVG KLLAKAQILLPIKEYIMERNPIVWEPLQPVV SRQALGHQAGESRGHTQRCKVTRLSSWQ VLVGAAVPCSGARDRVPVPRHVPQACLQG RVQTGRLDWRGHACASAPNAVPTVTFSDV AIDFSHEEWACLDSAQRDLYKDMVQNY ENLVSGLSITKPYVITLLEHGKEPVMVEK KLSKGMIPVLEVLARAMRQKNEIKGIQLG KEEVKLSLFADDMMIVYLENPVSAQNLLKL ISNFSKVSEIPKSMYKNHKAFLYTNNRQTE SQIMSELPFTIASKRIKYLGIQLTRDVKDLF KENYKPCSTK
2792	A	154	331	IPAAATCMGSLG*ETPGLWARRSVKSR GLFPGLPSPSRASVRSLLLPAAWAFLEGIV DTRPTAWRAFPWTLFLSVFCQFLDFPETS DSQKLSLDTPSF
2793	A	213	446	ILLQSLGVGGHRAWGIEPSKVLVSGRRT EAPSMQLMQGRQMWGRTSWRWTRTWRCG WPWGGPLAARHVSSCTKQGH
2794	A	515	278	IFTLFDKLSSQPSILRSQYQSCLYDPSQPWP PPTSDAHDHKGPHIAPPPPLCCLGLASPF RSIQYISARPQLKGPF
2795	A	1	708	VTAGVPKGHCPRRGTSIAIASCPYPGSPR AECALRAGSTVTT*RRSCCTS YSSGRPPTG RRGSWTLVCTSCCASWRRACSRSSSTSSS ATARCLRPWDSLRPCSGPSPSTSSGPSSCSE AFTVRHWTPSMRCSRMPQRSASIPYMS/S SDQPTPKS*RLQNVGSSS*DEGIPHVHTPG



Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GICQPCSGDKAGFRGSRAQPARKPSPTVQR KQNFNGKLVCFIPLGSAGKAVTWV
2796	A	2	590	FQGRGLAANDGEYLKLQWRAGTLVLAPS CPLSTLSVLSSPPRELQAMEALQNGQTTVE GSIEGQSAGAASHAMIEKILSEEPWQETA YVLGNYKTEPCKKPPRLCRQGYACPYH SKDRRRSPRKHKYRSLG/TQEASHGREEW QGRGQAEAAPTGSPGGGEAGPGDDRIASP GPRGGHSEDSWTVGAQLHLLHE
2797	A	319	513	IELRAVAQGIAQSLGQLLFTQCPLKDDLE GLFLQNNKEGVQKGRDEPLPLP*ATALSS IQAGIQAR*EGDLEAWQFPVRIHPPDQGG NIIVTFEPFFKLFKEFKQAVNQYGPSPFV MGLLKNAVSSWMIPTDWDALTRACLT AQFLQFKTWWADEAGRV
2798	A	1	915	MSTAVVVKVVLCTVAPGRGSAPSLSSCLD WKVNGAEGSHNKDLFVLTYGALVAQLCK DYEKDEDVNQYLDKMGYGIGTRLVEDFL ARSCVGRCHSYSEIDIIAQDMERGFCALHI DTEGRYEWWTSTQLQSTLPRAAQCSVYQ KQPDRKSLTVGQKIEVGNPGIGTEQSPQGL VRFATQAFLTTHRAEGLQSSQVKGSVIHL KSQDKCGEHRFTTNQVETGDPVRESSQH SVGRGGPKDIQIGANVPVRQCNLLWRITL GPLETPHLEFSGECSLAAMEAPEHTWDQ EKSDIPEPPHRSS
2799	A	75	642	EKLLNPQTTSFFLQLLQKKQWYPKSFPCCL PSQGLLPAARVQKCLLVLRNVSGSPFFLI GFPPPILELKESYPWAGTDIQCEPAQGHVL TSPSPTLRLQGAPDLPAGEPAWLLLTAREE DDG*NFSC*ASLVVQGQRLMKTTVIQLHL CEWRPDLSCQNKDYFPIRELLGQQCFIIT VATFFSL
2800	A	1	1146	MVGECGKLEVMQVHLSNPRDELEGELRS IRVTMGQVWALVHSTLEPFHTNEEEGLY NKVTEEVTEQVCLPAKAKAAKEGEVHPYP SPFPHYFEETEWPDPDLSFLEDTGDPSTL SHWQLTKEAEAELQIEKQVHKAQINRIDP EKIPDLLIFSTQHSPTGVIVQEODLVEWFL PHTNSWTLTPYLDQATMIGNERTQIVKL HGYDPRKIIVLLMKANIQQAFINGLTWQTH LANFVVILDNHFPMKMLFQFLKLTNWILPK ITKFKPIKGAENVFTDGSSNGKASYSGSKG LSQQLIWISSRNLPYHESDAEEEEIPGRTQG TPGCSHVETDTEEDPNCHEQHPLNTATHL GTDQEA VTDGGRKPEERGTTSHNE
2801	A	2	926	RPEPSCRPRSEYQPSDAPFERETQYQKDFR AWPLPRRGDHPWIPKPVQISAASQASAPIL GAPKRRPQSQRWPVQAAAEAREQEAP GGAGGLAACKASGADERDTRRKAGPAW MVRRAEGLGHEQTPLPAAQAQVQATGPE

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				AGRGRAAADALNRQIREEVASAVSSSYRN EFRAWTDIKPVKPIKAKPQYKPPDDKMVH ETSYSAQFKGEASKPTTADNKVIDRRRIRS LYSEPFKEPPKVEKPSVQSSKPKKTSASHK PTRKAKDKQAVSGQAAKKKSAEGPSTTKP DDKEQSKEMNNKLAEAKE
2802	A	25	435	TKYWLLFFLLILPFFFWRRSRSVTQAGG QWHDLGSLQPPPPGFKQFSCSLSPSSWDYR RAPLHLANFYIFSRD/MDFTMLARLVNSNR SQ/CDPLASASQSA GISGKSQHTRPVLVLK TYTNH/SF*VKGLGWEFIL
2803	A	1186	1074	TAAARRSSRTSSHRSLHVPENLATGPSEF RSPGFLLSRVPSVWDPTENRTVQLTWQPLP EPLELWPKA/HLTDSFPDLLGLAED*HCPI ASEAP*TIIDAELRVTLTVEGKPPFPLINTE ATHSTLPSFQGPVSLASITVVGIDGQASKPL KTPQLWCQLRQYSFKHSFLVIPCPVPVLG *DTLTKLSASLTIPGLQLYLIAALLPNPKPPL RPPLVSPDLNPQV*DPHSCPPENKPLTVIF LYLPKSYKTAPPHLPLLTFSDSARLHPGEI NSHVAHTKPVWWSLHTDAHEIWCRRSDR GTSLGRSIPCPPALCSMRKIHLRPQVLRQTS PRNISPISNPVSGFLLSPTCLTIPQLSPFN LGATLQSLPSLNFSHFHFLVETKETRFICGP KTPALVTDWEGSLPLMFNHCRDTSIIHPC FQGVPRCDACLSPSPLAASPAFLGKGQVP LNPFFTLGKSRFSGGGASTPTPSFHVSTPS LLFWGRGKYPSTPSSPLVASPAFLGKGQVP LNPFSFTLGGKSHFPGTGARFN
2804	A	3	810	GVSPCWPGWSRTPDFGSNPKCPPIRASPGA ELQALSSTVTTPYWGILVTA VFPH*GLRPR QCRQDHPAGRQGPGEVPEILGQSGCTD RTWSKAGGRTQAPGPRSRAGRRVSGQEIR APGPLGCRHGG/VGAPWTPEAASPLTATEP SCPH/LQAPCGYMPLSVSPRRRYRGPAGDQ KVKMLKFKAFLDYWQFLCLQPLHGAYK RSDLMTWIWGLLPEVTGAAGTTSPNVHT SGRFFRACVFCPVHTLVKKEPHPGQQEIM EPSPWSP
2805	A	62	475	FEPLFYLMCLLNLFLQLPRHPFLFTVDLV NTWGCPLPSSPQ*EWLLAAPHIRSTPPPLSS GFPARRQLEPGAGARGP/HHTQALHLSFFF VFLRRSL/DSVAQAGVQWRGLGSLQPLPPG FVILSSPLSLPSLT
2806	A	3	4804	KRLNIQKTLEVAFSEAVWMQPSVVLLDD LDLIAGLPVPEHEHSPDAVQSQRALAHALN DMIKEFISMGSLVALIATSQSQSLHPLLVS AQGVHIFQCVQHIQPPNQEQRCIELCNVIK NKLDLDCDINKFTDLDLQHVAKETGGFVARD FTVLVDRAIHSRLSRQSISTREKLVLTTLDF QKALRGFLPASLRSVNLHKPRDLGWDKIG

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GLHEVRQILMDTIQLPAKYPELFANLPIRQ RTGILLYGPPGTGKTLTAGVIARESRMNFIS VKGPELLSKYIGASEQAVRDIFRAQAAPK CILFFDEFESIAPRRGHDNTGVTDRVVNQ LTQLDGVVEGLQGVVLAATSRPDLIDPALL RRGRLDKCVYCPPPDQDSSSSSDSLSLSS MVFLNHSSGSDSAGDGECGLDQSLVSLE MSEILPDESKFNMRYLYFGSSYESELNGT SSDLEDESMNQPGPIKTRLAISQSHLMTAL GHTRPSISEDDWKNFAELYESFQNPKRKN QSGTMFRPGQKFFDEITELTYLPSFHHKAA PHQAEPGNSSASAPPPYNPFTSSPHTQS GLQFRSVTSPPPSAQQFPLKEVAGAKGIVK TALETAPTLALPVSSQPFSLHTAEVQGCAY GILTQGPGPCPVAFLSKQLDLTVLGSPSCL HAVASAAALILLEALKITNYAQLTYSSHNF QNLFSFSLTHLSAPRLQLYSLFVESPTIT ILPGPDFNLASHIILDTPDPDDCMSLIYLT TPFPHISFFSVPHVDHIWFTDGSSTRPDRHS PAKAGYAIESSTSIEATALPPSTTSQQAELI ALTRAFTLAKGLHVNIYTDKYAFHILHHH AVIWAERGFLTQGGSSINASLIKTLKAAAL LPKEAGVTHCKGHQKASDPITLGNAAYADK DRTIDGSSQVIBKNGYGVSDTGTLEVA ELEKLPNNWSPQTCELFALSQALKYLQNG KTISILIQKEPSALGLTPERKGNVGHAGKG PLESSSPDPFLCGQERREKGCRTATSVSITN PINRGPWVVVTHPGKELTPEHKGNVGHAGR DILAKAGAIHNLNIGEGTPVCCPLLEGINPE VWATEGQYGRAKNARPVQVKLKDSTSFP YQRQYPLRPKAQQGLQKIVKDLKAQGLV KPCSNPCSTPILGVQKPNRQWRITLCHQAT QALFNFLATCGYMSKPKAQLCSQQ/RYL GLKLSKGTRALSEEHIQPILAYHPKTLKQL RGFLGVIGFCRKWIPRYGEIARSLNTLIKET QKANTHLVRWTEVEVAFQALTQAPVLSL PTGQDFSSYVTEKTGIALGVLTQIRGMSLQ PVAYLTKEIDVVAKVVAVALVSEAVKIIQ GRDLTVWTSHDVNGILTAKGDLWLSDNCL LLKQCALLLEGPVLRCTCATLNPATFLPD NEEKIKHNCQQVISQTYATRGLLEVLPLTD PDLNLYTDGSSFVEKGLRKVGAYVSDNG ILESNPLTPGTSAQLAELIALTWALELGEEK RANIYTDKYAYLVLAHAHAIWKEREFLT SERTPIKHQEAIRKLLAVQKPKEVAVLHC RGHQKGKEREIBENCQADIEAKRAARQDP PLEMLIKQLV
2807	A	1	591	MTPRGTTGGDSEVPFQAAPLSVKQGVSR LWARRRPRCDFLRSSRIRVHPTPAASTMPP KFDPNKIKVVYLRCTGGEVGATSALAPKIG PLCLSPKKNRQAQIEVVPASALIKALKEP

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				PRDRKKQKNKHSGNITFDEIVNIARQMRH RSLARELSGTIKBILGTAQSVGCNV DGRHP HDIIDDINSGAVECPAS
2808	A	1094	483	IGCDVLINNAGIFQCPYMKTEDGFEMQFGV NHLGHFLLTNLLGLLKSSAPSRIVVSSK LYKYGDINFDDLNSEQSYNKSFCYSRSKLA NILFTRELARRLEGTNVTNVNLHPGIVRTN LGRHNTFHCWSNHSSIW/WSWAFFKTPVE GAQTSIYLASSPEVEGVSGRYFGDCKEEL LPKAMDES VARKLWDISEVMVGLLK
2809	A	1775	1981	HIWQNSLIVLFRGCRSAHAKVHRWKN*LP LNLAPLLPRSGSSAPIRPPPSAQARQPMKST YGVDRRHS
2810	A	272	51	MLLLSSSLKCGTCQWQVQPAVAGSLEGG BEESMVSALLISALPFLGTSHVTVETLDVQ YTVFPKLCIFLPC*
2811	A	3	357	FGFNGCSKRIKLQELSDLEERENEDSMVPL PKQSLKFFCALEVVLPSDCRSPGIGLVEEP MDKVEEGPLSFLMKRKTAKQLAIQKALSD AFQKLLIVVLG/QDCLDHP*STSVSVSK
2812	A	94	3006	RTRSLTRKAMAEHAPRRCCLGWDFSTQQV KVVAVD AELNVFYEESVHFDRLPEFGHV LDVHGVHVHKDGLTVTSPVLMWVQALDII LEKMKASGFEBQSQVLALSGAGQQHGSYIW KAGAAQALTSLSPDLRLHQQLQDCFSISDC PVWMDSSTTAQCRQLEAAVGGAAQALSCL TGSRA YEFNLVCDRKHLDKDTTQSVFMAGL LVGTLMFGLCDRIGRKATILAQLLLFTLIG LATAFVPSFELYMALRFA\GLLPSLDLASA MSPY*QNGWGP HGGRRPWSWPSATSPSGR WCLRDSPTVSATGGSFRRPALRLAYCSSVLL LGSARICTLAPDPWEDGRGD TTDPENGLG Q*AETLPGAHEPAGPREDRPLRECPGSVQT PPAPEGDPDYLLCLVCGQSGVLRPEPPSGG LRPGRLSDAAHLWSC*GACPLFQHLHDAE VWPQVEP/RWGPWSWVA*CVSSSSSSQIC PWWSPCWLWWGKWPQLPLPSPMCTLPS FSPSSGRQAWGWWASSHGSGASSHHL*S CWESTTLSPCSSTAASPSWPA/SLCTLPE THGQGLKDTLQDLELGPHRSPKSVPEKE TEAKGRTSSPGVAFVSLGTSDTLFLWLQEP MPALEGHIFCNPVDSQH YMALLCFKNGSL MREKIRNESVRSWSDFSKALQSTEMGNG GNLGFYFDVMEITPEIIGRHRFNTENHKYF KGKGAPGHPMPSLKANFDLLACL RGVGSS TLLWP AVLGAQTRQAGVNEGRSQVADF LRIPVTGCPEQRRNPPSPAPLGTGGPAEER LQFPGVAGSRRGRGRILRAGGIGRASPEG TGAPRPRAGQGRGGPGKPESGGGGPVLR PGDCTCCVLKSQPRQRRGACSAMAFVRV LRVRQSVRPPRGVIVAALQRPETQGAPASS

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				ARPDGCPESRGGLALWRRRLRGYASRDRVL CNRRCPPHAARFSPKRTSPSGPHLHLMSSW AVP
2813	A	1	897	MTYGVGKGDMDVDGTKERGERIESALGTS HIMRVAEPQGSQSWCPDEELRPVGSPTA AQKLPSTPGALGPTHSTECCSIPLDKAQQ GLQKIVKDLKAQGLVKPCNSPCNTPILGVO KPNGQWRLVQDLRIINEALVPLYPVNPY TLLSQIPEEAWEFTVLGLKDDFFCIPVHPDS QFLFAFEEPSNPTSQLTWTVLPKGFDRSPH LFGQVLAQNLSQFSYLDLTVLRVYDDLLL AARSETLCHQATQALLNFLTTCGYKVSKEP KAQLCSQEVTYLGLKLSKGTRALSEERIQA ILA
2814	B	71	2167	XPAEALKDGEERQKNKKKAKKIKARMNF RAKEYESLMETKNSGSDSPYKAKLQRLAK DLLKQVQVQDSGSWANNKVSALDRTLGEI TRILEKENVADQIAFQAAGGLTALEHILQA VVPATNVNTVLRNSSMPQDSYMQCVTLCF AVTGRSYSIFDNNRQDPTGLTAALQATDL AGVLHMLYCVLFHGTILDPTASPKENYT QNTIQVAIQSLRFFNSFAALHLPFQSVGA EGLSLAFRHMASSLLGHCSQVSCESLLHEV IVCVGYFTVNHPDNQGDRAVRPPPHSAK SSASCPSSISVTHG
2815	A	1	473	EVRWNSPPTDSLSPDGGSELEFYLAPEPFS MPSLLGAPPYSGLGGVGDYAPLMVLMCR VCLEDKPIKPLPCKKAVCBCECLKVYLSAQ IQCPCTCFVWCFKCHSPWHEGVNCKEYKK GDKLLRHWASEIEHGQRNAQKCPKCKIHI QRTEGCDHM
2816	A	1	1286	RGAVFPGPEHSVPESVTFEDVAVVFTDEE WSRLVPIQRDLYKEVMLENYSIVSLGLPV PQPDVIFQLKRGDKPWMVDLHGSEEREWP ESVSLDWETKPEIHDAADKKSEGLRECLG RQSPKPKFEVHTPNGRMGTEKQSPSGETR KKSLSRDKGLRRRSALSREILTKERHQECS DCGKTFFDHSSLTRHQRTHTGEKPYDCRE CGKAFSHRSSLSRHLMSHTGESPYECVCS KAFFDRSSLTVHQRIHTGEKPFQCNCEGKA FFDRSSLTRHQRIHTGESPYECHQCGKAFS QKSILTRHQLIHTGRKPYECNECGKAFYGV SSLNRHQKAHAGDPYQCNCEGKAFFDRS SLTQHQLIHTGDKPYECSECGKAFSQRCL TRHQRVHTGEKPFECTVCGKVFSKSSVIQ HQRRYAKQGD
2817	A	94	255	MLYIECKSHKLVAPLAVFFALFFLLIFFWV AFSYPFELLFLQLRSRQADIGVQ*
2818	A	551	19	TGTIDKLQSGPHLLRDWAFHPPWRKICL HCKCPQEEHMTVMPLMEKTISKLMFDF QRNSTSDDDSGCALEEYAWVPPGLKPEQV

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				HQYYSCLPPEKVPYVNSPGEKLRKQLLHQ LPPHDNEVRYCNSLDEEEKRELKLFSSQRK RENLRGRGNVRPFPVTMTGAICEQVSMDSG Y
2819	A	236	559	MWLEPMQMGFLHMEKMAARTSAILD*G TLK*FHFTLTTLKALSSHTPIFPGTGELQLP VSPSVCLDQGMQLKPSTSSHLLKTVKPRM KRQSLHMKQSFEPKIYL
2820	C	209	592	METETKESGKNKKIPPKHQIENVGVGGLG AQDGLNQIGKIPVLSQSQRFGTMPAAFP CVFPPQSLQVSPQMSSKAWEKQSLPLPGLR GSPVERKNRNYDLCLPYCLRNIFNCRGKPV LFWRKANR
2821	A	381	55	PASLPCCSLISDCCASNQRDSVGVGPSEPGV GYSLVVRRFLSRSEKRNIRVGVTFRSRCV/L SPLSLTQKGNLTPCASQVRQCLALLRLAH GACTHWPAPT VWHS LVR
2822	C	2	166	MQKRHNCKKVHALPPAVLGFQRASGCRF ANKRSRITHFGGRRLSLTPASDSAGV
2823	A	164	423	RGPVSRNQPPFTRFPQTRKTTETHVRGQSL PRPGTQSLQTKAAQVPSQRLPKNPE*AV WLTQAPNAHPN*VARETPNCQTKSSTR
2824	A	792	389	PTRPPLAQAPRAHLSAQKRLLLMKQKG VMNQPMAYAAPSHGQEQHPVGLPRTTG PMQSSVPPGSGGMVSGASAPGPGFLGSQP QAAIMKQMLIDQRAQLIEQQKQQLREQR QQQQQQQQILAEQVTCPLA
2825	B	1279	1479	MVPLCQVRVAGVRAGLALVSRTSPLAPNL AGVLGSGAPPPPPGSPCLRALRLPQQKS GPLRELLSAHGSKDGLVVKAPTHFYDHLF PRLFVLMKLF
2826	A	1	412	MKALLALPLLLLSTPPCAPQVSGIRGDAL ERFCLQQLDCDDIYAQGYQSDGVYLIYPS GPSVPVPVFCDMTTEGGKWTVFQKRFNGS VSFFRGWNDYKLGFGRADGEYWLGLQNM HLLTLKQKYELRVDLEDEN
2827	A	3	711	KIADFGFSNLFTPGQLLKTWCGSPPYAAPE LFEGKEYDGPKVDIWSLGVVLYVLVCGAL PFDGSTLQNLRARVLSGKFRIPPFMSTECE HLIRHMLVLDPNKRLSMEQICKHKWMKL GDADPNFDRLIAECQQLKBERQVDPLNED VLLAMEDMGLDKEQTLQSLRSDAYDHYS AIYSLLCDRHKRHKTLRLGALPSMPRALGL SSTSQYPAEQAGTAMNISVPQVQLINPENQ IV
2828	A	1350	2203	TWRLDPQIISPKPQPGGTYTLEVVKSSKSK KVLSPHP*WPLRLWQR/GGSPEGGTQAPD GSLPPPPRPKSERVSGPKLSSGGR/EGSHP GGPPHITHP/DGEEKAKSSWFLREAKDPT QKPSHPVKPLSAAPVEGSPDRKQSRSLSI ALSSGLEKLTVTSGSIQVPTQAPQAGQM

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				VDTKRLKDSAVLDQSAKYHYHLTHDELISL LLQRERELSQRDEHVQELESYIDRLLVRIM ETSPTLLQIPPGPPK
2829	A	2	259	WQGGILGSDPTPPLTSPNLLQTACFREERD VRRERGQPLGDHSALCLPRRGVPVPCDGL LCWWGPPDAAEPLRGPSPARAGPVLP
2830	A	1	1062	MTADAVLIKNGSKDADWEYEEGDKLEBFL RSLNSSKPLYLGQTGLGNIEELGKLGLEPG ENFCMGGPGMIFSRVLRMRVPHIGECLE MYTTHEDEVGRCVRRFGGTQCVWSYEG RCSFRVVPDSAIEFSMDFEKILMLDPTLHPL CQNLLQRLNTMWKPPNVGLVPSKATAQA VRWSSLAMARAGAATMPGALSQGCIEVS RLLKKLPDDEGITMDTVGFAPLCWQLT LANHQRYFADGPQVCNMQPAPHFAS MRSSAASPTSLPAFADPAVPPLEHVYVW TLLLCQRWCTYMYMDSTATTLTKHCCCPP PIPIGVLLPADWGHIGPSSDSRSENKAMGS SPST
2831	A	2	238	TKLNPKIMDVGWPELHAPPLDKMCTICKA QESWLNSNLQHVVVIHCRGGKGRIGVVISS YMHFTNV SAR*DEDVSSLS
2832	A	3	162	RLHTANLGDSGFLVVRGGEVVRHSDQQH YFNTPFQLSIAPPEAGVVLSDR
2833	A	1	988	MPAEFFQRCSVIMVQLPWKEAHVERPHGE RDYTPDLQPDMEKFPGLRRALRPVVKTL LVQLEYRQAEKCEKRDWPSLPDYIFLLCW MLPALEYRTPSSSVLELRALRAPQPADSL LWDLVIVPITSLKSWQTPRGEVEGVTHEEI CASLKSALAVALLSMSDLTVGTPVTQPQTL NTMGIISRGGRGQVAALNRQRQVPELIIGI DILSSWQNPHGSLNGRGYINSLALCHNLIR RDLDRFLLPQDITLVHYIDHIMRLDSVKDK WLHLAPPTTKKEAQCLVGLFGFWRQHISH LETAL/RPVTGLWWKLN*LWAIKSPCNLN CLS
2834	A	4061	2827	EAGPAPLSAAAPGAGRGWPRPLAERRKGR GRRQPLRARLNRNRWAAGQGSTVQAATF GPAMAAAPLKVCIVGSGNWGS AVAKIIGN NVKKLQKFASTVKMWVFEETVNGRKLTDI INNDHENVKYLPGHKLPENVVAMSNLSEA VQDADLLVFVIPHQFTHRICDEITGRVPKKA LGITLIKGIDEGPEGLKLISDIIEKMGIDISV LMGANIANEVA AEKFCETTIGSKVMENGL LFKELLQTPNFRITVDDADTVELCGALKN IVAVGAGFCDGLRCGDNTKAAVIRLGLME MIAFARIFCKGQVSTATFLESCGVADLITTC YGGRRNRVAEAFARTGKTIEELEKEMLNG QKLQGPQTS AEVYRILKQKGLLDKFPFLTA VYQICYESRPVQEMLSCLQSHPEHT
2835	A	106	1814	QLLPTDTPGTGNSSPSLPHLPFAGACGLSIYN

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LVPTQQKRNPSSGSSGFILSRICFTNYSVPVPS LQMFRLQLPPVNSEETSHYEIPLGRRVEL RYPLRQGTEATDGGQVCGNEDMLIRDRVRK TRGSAPPPAHNLAPTEVALEDVLRIFTSW RGVDGALEKGGTSCPARAQLPAEPEDPLF RCLRVSRKLDREVRGLGLPRQLQGVWSTT YPRRHAIAEHAGSPKPLRKREPETWQANK KGVIGIQLVVTMVMASVMQKIIPHYSLAR WLLCNGRKYNHIESKPLTPKIDIDLHLET KSVTEVDTLALHYFPEYQWLVDFTVAATV VYLVTEVYYNFMKPTQEMNISLVCKVLFS LTHYFKVEDGGERSVCVTFGFFFFVKAM AVLIVTENYLEFGLETGFTNFSDSAMQFLE KQGLSQTLHLHINFLAPLFMVLLWVKPITK DYMNPPLGKESIPLMTEATFDTLRLWLIL LCALRLAMMRSHLQAYLNLAQKCVDDQM KKEAGRISTVELQKMVARVFYILCVIALQ YVAPLVMLLHTTLLKTLGNHWSGYLSRI YLYLTSG
2836	A	2	774	HSYSHSHGHGCGSPAGDTEQGYKPVWPVCS LFPDGSHPGV*QPIHEPA/QGRGGLPPWGA A*TPRAWRLA*RPRG*AALPWA*TSPGRPA SAPLAHTGSGCPSRPTRAPGSP/IPQNIKR PYPGEAFVPSRAG/PTVGVTFSHLAPSLPP FPSS*LSPSLPRTTTSCTRAILTPSS*QKLLY PPSRP/VVLLVRRARPPAAAPTSEEPERSP WETPHAAPSQLHELHETHSVAQKSDLLPA PEAM*PGSVSSRFLLY
2837	A	2	521	CSAAWAPKLQLLSVCRQQLPGNPRARSHS HHRRTTRARCPSCGQARHSAGSWHKLQFP LCPWKMRSPKMRSLKMPSES RMVVT LISALESTEQYHGGVYTPCDIDSNILSPDI SNNITEGVYTPCDIDRHLIPFFLPLDMRLQV LMPLDSGTCTSGFPEALRPSASD
2838	A	14	1256	WPCGAAPGLTHASERMFTLTMTIQAAPV MGWDRKPLKMFSEEMRGHLHHHKKCLT KILKVEGQVPDLPSCLPLTDNTRMLASILN MLYDDLRCDFERDHFRCICEEYITGKFDPO DMDKNLNAIQTVSGILQGPFDLGNQLLGL KGVMMEMVALCGSERETDQLVAVEALIH ASTKLSRATFIITNGVSLKQIYKTTKNEKI KIRTLVGLCKLGSAGGTDYGLRQFAEGSTE KLAQCRKWLCNMSIDTRTRRWAVEGLA YLTLDADVKKDFVQDVPALQAMFELAKT SDKTILYSVATTLVNCTNSYDVKEVIPELV QLAKFSKQHVPEEHPKDKKDFIDMRVKRL LKAGVISALACMVKADSAILTDQTKELLA RVFLALCDNPKDRGTTVAQGGGKALPLAL EGTD
2839	A	1913	1582	EDSGRLRLWICLSLSLSP*NRVSLCHPGWS AVARPQLTAARPSRLQQSSHLSLQSTWDH



Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				RHTPPYLALFFIFLFLVDMSTFTVPRPVLNS WAQAILPFRPLKVLGLLA
2840	A	44	376	MYMLLQAFWLWQETLTKTILLYKFTKPPAN TPVLGVNAQVCHSCLAALRIRKVNHGKRN FKAQPPNGKLPVLGCLCLLTDLIHALGYD CRDFPVSLEYAELVFLFVAY*
2841	A	522	693	LDFFLVFLQQFLPRPSSSEI*MLPGFPAAAY GPVAAAAVAAAARGSGRKVYGTGDSQA
2842	A	87	439	KTWTPQPRHPPHPETSKPTPPC*GPVLCSC LKVMRPLPP/PP*DLCSPELLAPGPRRSAG GCWACQRRKKMSCLGGAGVCLKQGHG MGLCYDLGLSTLAEPGSSARRLPARSAL
2843	A	1	409	MAETAVINHKKRKNSPRIVQSNDLTEAAY SLSRDQKRMLYLFVDQIRKSDGTLQEHDI CEIHVAKYAEIFGLTSAEASKDIRQALKSFA GKEVVFYRPEKDAGDEKGYESFPWFIKHS TNITSLSLWFFSSCTH
2844	A	1	894	MPGPMSLWLLLLVPLSLEHSDLRICFPQG VVSMESSSTGFIWTDVRAWQTSNRHVSSW REPRHSRMPPGAGLMERIQAIQNVSDIAV KVDQILRHSLLLHSHKVSSEGRDQCEAPSDP KFPDCSGKVEWMRARWTSDFCYAFFGVD GTECSFLIYLSEVEWFCEPLPWRNQTAQR APKPLPKVQAVFRSNLSHLLDMGSGKES LIFMKKRTKRLTAQWALAAQRLAQKLGA TQRDQKQILVHIGFLTEESGDVFSRVLKG GPLGEMVQWADILTALYVLGHGLRVTVSL KELQR
2845	A	2	1841	TNDKNHMITSVDEGEKAFDKIQPFMLKTL NKLVLVLARAIRQEKGIKGIQLGKEEVKL SLFADDMIVYLENPVSAQNLLKLISNFK VSGYKINVQKSQAFVYTNNRQTESQIMSEL PFTIASKRIKYLGIQLTRDVKDFKENYKPL LNEIKEDTNKWKKIPCSWVGRINIVKMAIL PKVIYRFNAIPKLPMTFFTKLEKTTLKFIW NQKRAHIAKTILSQKNKAGSIALPDFKLYW KATVTKTAWYWYQNRDIDQWNRIEPSEIP HIYNHLIFDKPDKNKKWGKDSLNFNKCW ENWLAICRKLKLPFLTPYTKINSRWIKDL NVRPKTIKLEENLGNTIQAMGMGKDFMT ETPKAMATKAKIDKWDLIKLSFCTAKET TIRVNRQPTWEKIFTYPSDKGLISRIYNEL KQINKKKSNNPINKWAKDMNRRFSKEDIY AANRHMKKCSSSLAIREMQIKTTMRYHLT PVRMAIHKSGNNRCWRACGEIGTVGYKN DRQETQRTRKLHNLEDKPYGEINQIFLQV GQRKNGYARPQKSCLPCNIFQYVFQKKMK EKTKKEKKWNLGNTRIKPEKGKENMGGT VLPSSPIIWVEYEPVSSP
2846	A	60	493	EAGKRESSRDKGARCYYTRHGLRASIPAP GLRSRRGEQGCGRPCGKRLVCPGCRNQ

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				ENPEGNRGKGAARFTRESASGRGESRSAR GSIERSGDMRTYWLHSVWVLGFFLSLFSL QGLPVRSDVFNRTDNITVRQGDITAIL
2847	A	395	3	GGQGVTPWPSSCLPGTGSPAPSPTRLGPT PRDRAEAIVGPDSATCSQTEGAQEGGRCLP PG/MELPAGDGAGRRVGQGGPEGQLGGQQ RGKGAGPQPPPEQPGLAWVGDRLIHPRL CLPPTCGHRAGSPGW
2848	A	514	738	MNSLSWGAANAVLLLLLLAWASPTFISINR GVRVMKGPSAFLSGDDMKFAIPKEKDACC IRESSTRXXRSGSAGL
2849	A	2	427	HVIKVLHDDWIFTPFIQGP*SM/CSSKNESR HIGS*RVTG*LLEVLKSL*SFGRNLALNM KSL/TSEVQEE*RKLNKTHRVQRDFDKDRK LAVGQSESPGHPTSEKPPSTSSSAGCMLCS LHISRGFQLRRKRLNGKCCPIQ
2850	A	3	409	RQEGEDSAGSWHSQGPQCQGRAGAGSG P*/GPATGLGLGQ*QDQSQKGQSSARPG *GQAFQGGQGRTRARSEAGKGQGDORS RAGP*HGQGLR*GKGRARAR*GSGPRPG* GQGGKYGRTRGNAAKAGPGLT
2851	A	174	446	MWLLPALLLLCLSGCLSLKGPGSVTGTAG DSLTVWCQYESMYKGYNKYWCRCQYDT SCESIVETTGEKGGKEWPRVHQRPFGSR LHCDH
2852	A	1008	1246	INNLSWQDYGESP*ALSNQTS*VVPILRPFP VFLLLLHLVVFQFIQNRQAITNHSI*QMFL TTPQYHPLPQDLPSA
2853	B	428	3792	MSFDPNLLHNNHNGHNGYPNGTSAALRETGV IEKLLTSYGFQICSERQARLFFHCSQYNGNL QDLKVGDDVEFEVSSDRRTGKPIAVKLVKI KQBILPEERMNGQVVCAPHNLESKSPAA PGQSPTGSVCYERNGEVLYLTYTPEDVEG NVQLETGDKINFVIDNNKHTGAVSARNIM LLKKKQARCQGVVCMKEAFGFIERGDV VKEIFFHYSEFKGDLETLPQDGVFTIKD RNGKEVATDVRLPQGTIVFEDISIEHFEGT VTKVIPKVPKSNQNDPLPGRIKVDVFKEL PFGDKDTKSKVTLEGDHVRFNISTDRRDK LERATNIEVLNFTQFTNEAREMGVIAAMR DGFGRKCVDRDVRMFFHFSEILDGNQLHI ADEVEFTVVPDMLSAQRNHAIKLPKGT VSFHSRSDHRLGTVEKEATFSPKTTSPN KGKEKEAEDGILAYDDCGVKLTIAFQAKD VEGSTSPQIGDKVEFSISKQRPQQVATC VRLGRNSNSKRLGYVATLKDNFGFIETA NHDKEIFFHYSEFGDVSLELGDMVEYSL SKGKGNKVSAEKVNKTHSVNGITEADPTI YSGKVRPLRSVDPTQTEYQGMIEIVEEGD MKGEVYPFGIVGMANKGDCLQKGESVKF QLCVLGQNAQTMAYNITPLRRATVECVKD

500

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				QFGFINYEVGDSKKLFFHVKEVQDGIQLQA GDEVEFSVILNQRGTGKCSACNVWRVCEGP KAVAAPRPDRLVNRLKNITLDDASAPRLM VLRQPRGPDNSMGFGAERKIRQAGVIDXN WRKQKCFVFTKINGLFTQRSKPQTTRGKIK PPSPTSPELTLVILDKAFSPLARDPVYGGFK KRAKKS DPSIPVI
2854	A	1	747	MRLQRPRQAPAGGRRAPRGGRGSPYRPDP GRGARRLRRFQKGEGAPRADPPWAPLGT MALLALLLVVALPRVWTDANLTARQRP EDSQRTEGDNRVWCHVCERENTFECQNP RRCKWTEPYCVIAAVKIFPRFFMVAKQCS AGCAAMERPKEEKRFLLLEPMPPFFYLKC CKIRYCNL/GGA/NLSTHQ/CSKNMLGAWV RAVVGCGWPSSCCWPPLPASACLEPRDC HRLSLPEHGLAPDRCHLLH
2855	A	3	1018	FASFPSINLQQLKEVPKRFGDERGAIVHY TILNNHVYRSLGKYTDFKMFSEILLSLT RKVLLPDLEFYVNLGDWPLEHRKVNGTPS PIPIISWCGSLDSRDVVLPTYDITHSMLEAM RGVTNDLLSIQNGTGPSWINKTERAFFRGR DSREERLQLVQLSKENPQLLDA/WNYRIFL FPRERKGA*KAKLMGLLDTC*RNVDGTV AAYRYPYLMGLGDSLVLKQDSPYYEHFYM ALEPWKHYVPIKRNLSDLLEKVKWAKEN DEEAKKIAKEGQLMARDLLQPHRLYCYYY QVLQKYAERQSSKPEVRDGMELVPQPEDS TAICQCHRRKPSREEL
2856	A	3	3707	RAGEVVPGWLLAAAAAHPGRPAASLSPGL GAVLGVAGRQVADPRFRRDWFRIPSPAE SAGPARQAGFAAAPPARAGPALSTMKGTR AIGSVPERSPAGVDLSLTGLPPVSRRPGSA ATTKPIVRSVSVVTGSEQKRKVLEATGPGG SQAINNLRRSNSTTQVSQPRSGSPRTEPTD FLMLFEGSPSGKKRPASLSTAPSEKATWN VLDDQPRGFTLPSNARSSSALDSPAGPRRK ECTVALAPNFTANNRSNKGAVGNCVTTM VHNRYTPSERAPPLKSSNQAPSLNNIIKAA TCEGSESSGFGKLKPNVSSATHSARNNTGG STGLPRRKEVTEEEAERFIHQVNQAAVTIQ RWYRHQVQRRGAGAARLEHLLQAKREEQ RQRSGEGLTLLDHQKEAARRKAREEKAR QARRAAIQELQKKRALRAQKASTAERGPP ENPRETRVPGMRQPAQELSPTPGGTAHQA LKANNAGGGLPAAGPGDRCLPTSDSSPEP QQPPEDRTQDVLAQDAAGDNLEMMAPSR GSAKSRGPLEELLHTLQLEKEPDALPRPR THHRGRYAWASEVTTEDDASSLTADNLEK FGKLSAFPEPPEDGTLLEAKLQSIMSFLDE MEKSGDQLDSQQEGWVPEAGPGPLELGS EVSTSVMLRKLEVEEKKQAMLLLQRALAQ

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				QRDLTARRVKETEKALSRQLQRQKEA\YE ATIQRHLAFIDQLIEDKKVLSEKCEAVVAE LKQEDQRCTERVAQAQAQHELEIKKLKEL MSATEKARREKWISEKTKKIKEVTVRGLEP EIQKLIARHKQEVRRLLKSLHEALLQSDER ASQRCLRQAEELREQLEREKEALGQQERE RARQRFQQHLEQEQWALQQQRQLYSEV AEERERLGQQAARQRAELEELRQGLEESS ALTRALRAEFEGREEQERRHQMELNTLK QQLELERQAWAAGRTRKEEAWLLNREQE LREEIRKGRDKIEILVIHRLEADMALAKEE SEKAAESRIKRLRDKYAEELSELEQSERKL QERCSELKGQLGEAEGENLRLQGLVRQKE RALEDAQAVNEQLSSERSNLAQVIRQEFED RLAASEEETRQAKAELATLQARQGLELEE VHRRVKTALARKEEAVSSLRTQHKGSVK RADHLEELLKQHRRPTSTKCPGMPGTLFK NGRQRTKAGRGPRGPQGRPPAPHRGWL RCPRLSTCGCILTVEAVVFSKKKKKGAPF
2857	A	1	2064	MTASIRRYHTCATDGEPPSSVLVGGDGD TLLVAALGLDLGLPFMLLPMEWMRVAI TYAEHRRSLTVDSGDIRQAARLLLP/GPEH CFSSFR\RLDARAATEKFNQDLGFRMLNCG RTDLINQAIEALGPDGVNTMDDQGMTPLM YACAAGDEAMVQMLIDAGANLDIQVPSNS PRHPSIHPDSRHWTSLTFVLHGHISVVQL LLDAGAHVEGSAVNGGEDSYAETPLQLAS AAGNYELVSLLSRGADPLLSMLEAHGMG SSLHEDMNCFSHSAAHGHRGIWGLVTLGP LACLEEDHETPSRPVPQSSPSGQEGTGGQ LRNVLRKLLTQPQAKADVLSLEELAEGV EESDASSQSGSGEPVRLSRTRTKALQEAM YYSAEHGYVDITMELRALGVPWKLHIWIE SLRTSFSQSRYSVVQSLLRDFSSIREEEYNE ELVTEGLQLMFDILKTSKNDSVIQQLATIFT HCYGSSPIPSIPIRKTLPARLDPHFLNNKE MSDVTFLVEGKLFYAHKVLLVTASNRFKT LMTNKSEQDGDSSKTIEISDMKYHIFQMM MQYLYYGGTESMEIPTTDILELLSAASLFQ LDALQRHCEILCSQTLSESANVTYKYAKI HNAPELALFCEGFFLHKMKALLEQMPSGS SSTAAAAKCRAWIHCRCTCRTPWQSACTLS TSPPGSAA
2858	A	1	571	FRPGRRAKRAMAVYVGMRLRLGRLCAGSS GVLGARAALSRWQEARLQGVRFSSREV DRMVSTPIGGLSYVQGCTKKHLNSKTVGQ CLETTAQRVPEREALVVLHEDVRLTFAQL KEEVDKAASGLLSIGLCKGDRLGMWGPNS YAWVLMQLATAQAGIILVSVNPAYQAME LEYVLKKVGCKALVFPKQ
2859	A	2737	2600	MCCWTWFASILLRIFALMFIRDIGLKFSFFV

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				VSLPGFGIRMMLAS*
2860	A	1	1353	MVKLSIVLTPQFLSHDQGLTKELQQHVK SVTCPCYLRKVSLKTFWSRNGHDGSTD VQQRRAWRSNRRRQEGLRSICMHTKKRVSS FRGNKIGLKDVTILRRHVETKVRKIRK VTTKINHDKINGKRKTARKHTGDCHPGE VVGQAHFVPDSPVHIALHGMAQPLFGIQG GALEPAGRGTGFLDSPVFRPIRKYNVQIPPS ARKALCNWSLLLCVGVKPEIFVAIHYYTPN TKLVPLARPRNSHVPHPPERTTVTQYSTCA LLTALCLLLPVLQETAQSRRMVTSHPEDSP ALARKHGASQAGLGFPRTQTVTPAFTFQT PTAAEPALLSAWLGRAPETETTTDMAGSA AAAPTCEMLRAHGDDLYFKWEPACASSQ AITVLPKHS GTGGSRQGPVAHPAAPFPKV RGEGTYYLHLSVFSDDLVLHLLHVQGRV VQGLRLRL
2861	A	1553	1896	CSSFCEFPFRSRPTAPRDPHRAEPQRLHSA EGAPEVVGPTSDPHHHPCGGAPGGTQDP KMAAEAPQQPNSDWAGEISMCRGSTHQL QMAFSETFLSALSGSSRGRPAGKESC
2862	A	262	129	SGLFLFFFPFPLPLPLCKHQIRDEWGNQI WICPGCNKPDDGSPMIGCDDCDDWYHWP CVGIMTAPPEEMQWFCPKCANKKKDKKH KKRKHRAH*RDDYKMLFMTYKRKLRI FV RNALSLNT
2863	A	3	520	LVDPVRVAVFLQLPLLSRAQGNPGASLD GRPGDRVNLSCGVSHPIRVVWAPSFPAC KGLSKGRRPILWASSSGTPTVPPLQPFVGR LRLSDSGIRLELLLSAGDSGTFFCKGRHE DESRTVLHVLGDRTYCKAPGPTHGSVYPQ LLIPLGAGLVGLGALGLVWWLH
2864	A	1	553	RTRGRTRGLVIKKWASHHQINDASRGTLS YSLVLMVLHYLQTLPEPILPSLQKIYPSFS PAIQLHLVHQAPCNVPPYLSKNESNLGDL LGFLKYYATEFDWNSQMISVREAKAIPRPD GIEWRNKYICVEEPFDGTNTARAVHEKQK FDMKDQFLKSWHRLKNKRDLSILPVRA AVLKR
2865	A	516	848	MWSLWIWVDQHARLIPSPQVLLLLRET PSTAAAVAGWLVVASMLLQLHAVGGVA LTSSHPFMWATGEELRKPPWQGSAGSAG VEELTGKHSCPGPEEPATVQKAPA*
2866	A	349	1018	TFTQPD PDDLISKPPRTPGGG*YQTQWSPSP DPRRTSPAGRPGPARRPPRTPRPARGRHP GR*GGPGASRPGGTGAAPAADQTGSPAVS TPSEFGAPGQAEQPSPIRASARSHLSCTA WLGKPSKPSAQRQPTVGPDGRDGGSSQAP NLSRGQAWRASLSPQNTSATGRVTCHGQ STWPLCRLKSNRRRKSGFA/GNKSEPVGLT RRSKHQPRNPQGQVGI

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2867	A	117	560	MYTVSLLLCLEFFKKSDFDPGPFQNNLFHNH GTQSQSCMGSKVGDVIPGAARLISETAQRV HTIGQKQKNDQHLRRVQALLSGRQAKGLT SGRWFLRQGWLVVPTHGEP RPRMFFLFT DVLLMAKPRPTLHLLRSGTFACKALYPMA Q
2868	A	438	2	TQRLVISEPDGEILTPGWDTPQDRMGVESRT NIQELGNRNQREAGGENLPETQAHMGETQ DQLRCKIDAETQTPWENQDKNGSEDAVE TQTFEKKDKKEAGEEDGEEIQAQGLGKQG QTGDENGEETQTRVLRALETIPASS
2869	B	1	390	MTPKHDHLGHVLPISLQLLLELSSCLPAAS AVWCAGCNDPWMTGYPDNMHYNKPMML HDRGGSVTLASQSWYAGCNAEKSEVN AFPGTQGMRFISAASYKDWWVQLQKQDV SRNMGTKARSASSLKN
2870	A	1	3411	MMEGEGGV RMSHDQ TGNKRKHGTSGISV CPNLLLLQEYQPDYRAHASGLNLISSSKAL PKYSHVLSGLCKICSFGPRFSLHSDTFFAL FAHADPEQIRNCETPAPPLQTERKNEMRIK THPSSSPLYDTPGRPAGSDSSSRGRAGAL STFLBPQRPRTHLSLILHRPSPGPRLSLPLFT KPSFLGSGRREHAEERARGPRETA AVAAR AEQGRGGSHSHSALGAPRRVAMPLPGLAL LLLAAWTARALESLNRSAAAGGCRKEMN KGNDNGALAIGGNMVIIWVDDFGWYVDR DTLEQGSPTPSHGQVLVHGLLGTGPHSRST LNIKEQLPRSKISSIGACNIIFQVDINAIFGIL MVPTDGNAGLLAEPQIAMFCGRLNMHMN VQNGKWDSDPSGKTCIDTKEGILQYCQE VYPELQITNVVEANQPVTIQNWCKRGRKQ CKTHPHFVIPYRCLVGEFVSDALLVPDKCK FLHQERMDVCETHLHWHTVAKETCSEKST NLHDYGMILLPCGIDKFRGVEFVCCPLAEES DNVDSADAEEDSDVWVGADTDYADG RTSAIFGYDHDCKVHDAFALSSVLVDRQE WGSTYESGAGQGIAAFWGACWKEEQSL FLLPDMDWLCLHSNINFNYISQNSHMLWR DPGEIDSKKLSALSSLPGIVLALGKAQRILLI ELLGVGLESEDKVVEVAEEEEVAEVEEEE ADDEDEDGDEVEEEAEPEYEEATERTT SIATTTTTTTSVEEVVREVCSEQAETGPCR AMISRWYFDVTEGKCAPFFYGGCGGNRN NFDTEEYCMAVCGSATNCTFDLKKSWSSG GQIQMADSIQRKGAELEAICQKRFSQRKHR YGKCFVGVLPVMBEHEFVIGTLGAASPFM NKLKANLCYFTPENRALAVPTTAASTPDA VDKYLETPGDENEHAHFQKAKERLEAKHR ERMSQVMREWEAEERQAKNLPKADKKAV IQHFQEKVESLEQEAANERQQLVETHMAR VEAMLNDRRLALENYTTALQAVPPRVGL

504

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				AAAEFTLQVTAQTPRHVFNMLKKYVRAE QKDRQHTLKHFEHVRMVDPKKAAQIRSQ VMTHLRVTYERMNQSLSLLYNVPAAVEEI QDEVAFKINKNMNYYKPDAGKISG
2871	A	18	382	GKMPPHLAMGCPPRLNPWEQPELGARGR GDGCPCPAEHGWALDVRYSLPLPQSLASS LAIPPQVFCSTLSSKSPRPAARQETPAGAP PAGPSFAGRRRTIPGSGAPRRSPGGRRREQ LR
2872	A	673	941	CCLAAHSGPPAQGQRRGPG*LCCSAGSGG NL*S*AGGPG*GRSGQVCPWPWPGPAGPH RPALPGSGGSSAVGRSAVPGAVRSPSHAG W
2873	A	227	712	ALLESLSGGEAQAWGAPRLVAGIRLIEHKC VLGGGTAGAWG*KDQVTIQPAGHAPGLSG TEATVTPDDSVSDPTTWPSQEVSMCHPLPG SHPSHLLKEGMTSVRPRALQGGPPWQLQT KDSAPPP*TPASFSPFFLSPLPVSPSLSHTH SFRVQGAKRFA
2874	A	1942	932	ARVRWRPPRWPPRASCPCPALRLCRGGSM GGPRGAGWVAAGLLGAGACYCIYRLTR GRRRGDRELGRSSKSAEDLTDGSYDDVL NAEQLQKLLYLLESTEDPVIERALITLGNN AAFSVNQAIRELGGIPTVANKINHSNQSKE KALNALNNLSNVENQIKIKIYISQVCEDV FSGPLNSAVQLAGLTLLTNMTVTNDHQHM LHSYITDLFQVLLTGNGNTKVQVLKLLNL SENPAMTEGLLRAQVDSSFLSYDSHVAK EILLRVLTFLQNIKNCLKIEGHLAVQPTFTE GSLFFLLHGECAQKIRALVDHDAEVKE KVVTIIPKI
2875	C	1	531	MARNECVDGQPGHLVDFTCLVTYRVSGES RAPHMAELFLVYHMBEKELETHIPRKQER VEEKGPCICKALSPNSVNQRDAREKEMLQ QLQNRDTKQVLPKASAHPLDKAHHTAK PDGSGGEKDFLHTRTPPPLLQGRAGNIFN NKTVYRSNTIITIGRWVLRALRPKDNN
2876	A	1573	2858	EPVFEQAIDQRSSTDTSLSTPAAPMVDSLIA RVGVMARGNAITLPVCGRDVKFTLEVLRG DSVEKTSRVWSGNERDQELLTEDALDDLIP SFLLTGQQTAFGRRVSGVIEADGSRRRK AAALTESDYRVLVGELDDEQMAALSRLG NDYRPTSAYERGQRYASRLQNEFAGNISA LADAENISQ*ICWKYFCAG*CGKYFRKIIT RCINTAKLPKSVVALFSHPGELSARSGDAL QKAFTDKEELLKQQASNLHEQKAGVISP PEEVITLLTSEIKTSSASRTSLSSRHQFAPGA TVLYKGDKMFTTVKIAKRSQAPCMKSNNA LIVILGTVTLDVAVGIGLVMPVLPGLLRDIVH SDSIASHYGVLLALYALMQFLCAPVLGALS DRFGRRPVLLASLLGATIDYAIMATTPVLW

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				IYPLVNSPSC
2877	B	448	3506	XALMIEIDGGESWSFMDDNQNKTHDKKE KKMVVQKPHGTMEYTAGNQDTLNSIALK FNITPNKLVELNKLFTHTIVPGQVLFVPDA NSPSTLRLSSSSPGATVSPSSDAEYDKLP DADLARKALKPIERVLSSTSEEDPGVVKF LKMNCRYFTDGKGVVGGVMIVTPNNIMF DPHKSDPLVIENGCEEYGLICPMEEVVSIAL YNDISHMKIKDALPSPGEWEDLASEKDINP FSKFKSINKEKRQQNGEKIMTSDSRPIVPLE KSTGHTPTKPSGSSVSEKLLKLDSSRETS GSPTVTKLKSEPSDTSSAFESTAKENFLGED DDFVDLEELSSQTGGGMHKKDTLKECLSL DPEERKKAESQINNSAVEMQVQSALAFGL TENDVELKGALDLETCEKQDIMPEVDKQS GSPESRVENTLNHEDLDKVKLIEYYLTKN KEGPQVSENLOKTELSDGKSIIEPGGIDITLS SSLSQAGDPITEGNKEPDKTWVKKGEPLPV KLNSSTEANVIKEALDSSLESTLDNSCQGA QMDNKSEVQLWLLKRIQVPIEDILPSKEEK SKTPPMFLCIKVGKPMRKSFAHTAAMVQ QYGKRRKQPEYWFAVPRERVDHLYTFFV QWSPDVYGDADKEQGFVVVEKEELNMD NFFSEPTTKSWEIITVEEAKRRKSTCSYYED EDEEVLPVLRPHSALLENMHIEQLARRLPC KGYPWRLAYSTLEHGTSLKTLYRKSASLD SPVLLVIKMDMDNQIFGAYATHPFKFSHDYY GTGETFLYTFSPHFKVFKWSGENSYFINGD ISSLELGGGGGRFGLWLDADLYHGRSNSC STFNNDILSKKEDFIVQDLEVWAFD
2878	A	226	2263	SVKNYTKCHVRNEQICNKLTSCKSCSLNL NCQWDQRQQECQALPAHLCEGEGWSHIGD ACLRVNSSRENYDNAKLYCYNLSGNLASL TTSKEVEFVLDEIQKYTQQKVSPWVGLRKI NISYWGWEDMSPFTNTTLQWLPGEPNDSG FCAYLERAAVAGLKANPCTSMANGLVCE KPVVSPNQNARPCCKPCSLRTSCSNCTSN MECMWCSSTKRCVDSNAYIISFPYGCLE WQTATCSPQNCGLRTCGQCLEQPGCGW CNDPSNTGRGHCIEGSSRGPMKLIGMHNN EMVLDTNLCPKEKNYEWFSIQCPACQCNG HSTCINNVCQCKNLTTGKQCQDCMPGY YGDPTNGGQCTACTCSGHANICHLHTGKC FCTTKGIKGDQCQLCDSENRYVGNPLRGT CYYSLLIDYQFTFSLQEDDRHHTAINFIAN PEQSNKNLDISINASNFNITWSVGSTA GTISGEETSIVSKNNIKEYRDSFSYEKFNFR SNPNITFYVYVSNFSPWIKIQAIFSQHNTIM DLVQFFVTFFSCFLSLLVAAVVWKIKQTC WASRRREQLLRERQQMASRPFASVDVALE VGAEQTEFLRGPLEGAPKPIAIEPCAGNRA



Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				AVLTVFLCLPRGSSGAPPPGQSGLAIASALI DISQQKASDSKDKTSGVRNRKHLSTRQGT CV
2879	A	1	1131	MKVTFANKPEGGGRLAKQRPPGRGARPRP KHEGGQSVLGTTRRPALLQVSCTDVSLEQ DKDGATATHFAASRGHSKVLWLLHGG EISADLWGGTALYDAAENGELGCCQLVV NGAELEVRDRDGYAAADLSDFNGHSHCT HCLRTVENLHRGMVLALGAAEHSKAQR EAAGGPEGELPEKESLEENWPSRGQGLV PSAPTAVAQSMHCVLSDPSVELEAKQP DSGMSSPNTTVSVQPLNFDLSSPTSTLSNY DSCSSSHSSIKGQHPPRAPNPQILQYKKRFS ELEQLLERSGELEQQQLRDAEHSQDLESAL IWLEEEQQGGPGLAAWPPGRAPTDPLCPIQ ECQPGPGECALRTAGPGRFGQPGSE
2880	A	1	416	FRTDARVAITYYQATEEFQNGIASYIPKDN SLQSETVQYKRGVCQQFCLPSHTVDPSEW AEBELGFDLDREVYPLVVHAVVDEGDEYF GHCHVLLGTFEKHTDGTFCVKPLKQKQVV DGVSYLLQEIYGIENKYNTQ
2881	A	419	1	KYFKCAPFPATRPKAHTVFLKNVDIQVNL RFCSKVAKLHYPNNLLFHSLGITKMQDR KELAVVQSHSGSKGRILFSPSLPALEQLRVP LEEHSASPDPIHPPSLAPERAAASGPPTGAE TRVPAPHAGTDPSEPFR
2882	A	2	366	ARPRVVLKRLGSQRELAQLGPEHLQAGHR PAPLRPAAGHAPDRVRAPQRRRASAHARG SGGLVGPALPLAAPSRRPGAPLRGDQGL GQLPASQPQGLGAHAAAADPGLQPRAAG ATEFSV
2883	A	3	1396	RQENNTRGVPSLLKSFLQERLGIHLIRKIV KPKHHVLMRSKESWKVKSEIPKVPKQPLV LHHPRMTTTKSPSKDMLEPEAELAEDLPTT KSTSVES/EDAH*EPGRFPFVLPDL/PCHCLP SAPTPLCIVKRPCPT*VTQLSASAQSAHQ RTPRAQSPSS*PR*VNCLPPS/LHKDDLELK EKDQKKPPTAPREVKGTRRKLPTAFLPSKY HGYEELLTAKPDPAFTEPKGIQKNA/PSPAT NAEAPTPVPLLQAQAGHSSETLCSQRETGP ENPDSTPKED*SPTSG*HLHSLAGSPEHYRG STRCCPAPVDRTAAGEP/ASSTWRPRGC*R SSRHVTGSW*VALCAQCSGLPRSPWPAQR *VRASPSSATSSSSWMSSARSPPVTHKAR AVHGGCVHHPACAPALPEGSVPWTAPQG* PAGHRPQSSAGPHLLATRWHLVRISSPPWP RHDLPVGPAAIKSGCTGQ
2884	A	437	748	MLIGLLAWLQTVPAHGCQFLPITSVTATVY HLPVHQLKGRSRVQKNLTLDNEGEGTWT CLEFLESLAGWRLGWGVSRGVREWLCLQ QVSLHQTPGLPHKQDL*

507

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2885	A	1696	2394	ERSTYDLRSSDRPAQETSHQFQIHLPCVLLL YSPTLTLKYISTPSLATDHAPLTISLKPNNP YPAQCQYPIQHALKGLKPAITRLLQHGLL KPINSYPNSPILPVLEPEKIYRLVQDLRLINQ IVLPIHPVVPNPYTLSSIPPSTIHYSVLDLK RAFFTIPLYPSSQPLFAFTWTDPTLQAQOI TWAVLPQSFTDSPHYFSQAQISSLSVTYLSI ILIKTHILLSLIMS
2886	A	377	3	TPAWMTERDCTWRRRTSAPGGSWPSGPVP SPGAQ*RPSPQGLGLWWAAAAAPRC*TAP GPRPPPHGPGSPQGASPTTRPPRCRPHRA GSAGPTGATPPGSTQGQRRRHSHQLPGHP GHRVALG
2887	A	1162	536	HILRRQEFFFFCLFVCLRWVLLVLLPRLE*CG MILAHCNLFLLGSSNSPASAS*VAGTTGVR HHAWIIFCILVETEFHRVAQTDELLSSGNP PASAS*SAGIIGVSHSAWPESCRYARRKCF CVKKLRRWKLNLPCIQKAVSEGHWCWQASP YRDSAVREQSIWGTASSGGARMRWSSPA ALYVRLLAGFSFINKLVASEYRVFSSTL
2888	A	128	2626	NSHRWVYVRARRWRRRGKQREQPEDRGV PMKRAAMALHSPQYIFGDFSPDEFNQFFVT PRSSVELPPYSGTVLCGTQAVDKLPDQOEY QRIEFGVDEVIEPSDTLPRTPSYSSSTLNQ APEFILGCTASKITPDGITKEASYGSIDCQYP GSALALDGSSNVEAEVLENDGVSGGLGQR ERKKKKKRPPGYYSYLDKGGDDSSISTEAL VNGHANSAPNSVSAEDADEFMGDMPPSVT PRTCNSPQNSTDVSDIVPDSFPFGALGSDT RTAGQPEGGPAGADFGQSCFPAEAGRDTLS RTAGAQCVCVGTDTTENLGVANGQILESSG EGTATNGVELHTTESIDLPTKPESASPPAD GTGSASGTLVPSQPKSWASLFHDSKPSSSS PVAYVETKYSPPAISPLVSEKQVEVKEGLV PVSEDPVAIKIAELLENVTLIHKPVSLQPRG LINKGNWCYNATLQALVACPPMYHLMKF IPLYSKVQRPCTSTPMIDSFVRLMNEFTNM PVPPKPRQALGDKIVRDIRPGAFAFEPTYTYR LLTVNKSSLSEKGRQEDAEYLGFI LNGLH EEMLNLLKKLLSPSNEKLITISNGPKNHSVNE EEQEEQEGESEDEWEQVGPRNKTSVTRQA DFVQTPITGIFGGHIRSVVYQSSKESATLQ PFFTLQLDIQSDKIRTVQDALES LVARESVQ GYTTKTKQEVEISRRVTLEKLPPVLVLHLK RFVYEKTGGCQKLIKNI EYPVDLEISKELLS PGVKNNFKCHR TYRLFV VYHHGNSAT GGHYTTDV FQIGLNGWLRIDDQTVKVINQ YQVVKPTAERTAYLLYRRVDLL
2889	A	1669	1338	FRPRRANKFRSRIRNQPGPHGETPFFL*IP KLARHGGG/CP*SPLLRVRPENPFNPGSRG FN*LKPQPCPPTWVTE*DSVSKTNKQPPT

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				KKNRDGRWGAIWESQMETWS
2890	A	807	369	GKGGCGQTRRCARPGRRHHAAPALRADRT GPAPRRGLFGRCTLQPSARRLSSEHSV*Q THGCATPSRCHGGDGRDRGSPGDRGERP AGPAGGAGLEPAPGTLQPRSRPSRRWLLSP GAGAAQLEVVHLPGQRPQNPQCPLDFLP
2891	A	1204	2	FFFFVPPPLFTDPRAPQPHRHAFRGHRKE KGPGDPPSTPQSQVADPAAAPQGQPGC/RLP RGHCDRRHQEARPGCWGPPGGPGSILGPK SWCHLEADSGKRPGWTVGVGVRRSSPACP GH/VEQQGSAGSPGWMGWGCPVVS*PLQ GQNQPSSSLGGSRGSSFFSPDPDPA/GGQQQE GEGRGERSGQGPWGPFSFKNA/RQVAGGG QEGGQGPDPHDGGSRLRPPRMKEGGLGRRG RPQPSVTPVLGSAARWSKAPPSQGDHRT GGNRHLAP*SSGGRGGAPGALGL/PWHPA CSGASGHSGRWA*RSSGWG*GPSHTPPPG PARHPAPGLAGLAPHPARLRK*SGRSPR/E AGVKISLLGGERGL/PGPLAVVHDSGDGG AGHRGGV*S*RSVPPDPLSLSPRPAA
2892	B	74	325	SAFSYIPRRLDPTESYYYRPAREQERPA GVLTSVVYGKRINQPIELNRDFGRANHVQ ADFYRKNDIPSLKEPGFGHAPS
2893	A	1	3426	MAGGQVEVAVVADQLCAKYSKEYGKLCR TNQIGTVNDRMLMHKLSVEAPPKILVERYLI EIAKNYNVPYEPDSVVMVEDILEMSLVEFG NIGEAFLQEQNSPESSVTLTSANATLLSRQ NISTLPLSSYTLGHPAPVRLGFPSALALKEL LNKHPGVNVQVFALDPVLGTFILTSVILM VLVVINLFVSAILMAFGKERKSLKVVMS NTICYRENRISTVPPSGTRETARKAKGHRG LPENPVQLSEAFNCQDKLCNWIPVGQCPA ARSTVYANERAQLPGVTMASRVIFPLPLA FESLHTPGKSSSQSDAGAGPPILGLFCPW TRGPRLSALRARRLSSPIADVKNKNIPPSKHR TILSSRPDGSILFLPPFFVVTITPPARADVQE KDGHTIEQDEGERQHQIEKTEENTNPKPKR KQKLAPGTPQSNMKPVHERSQECLPCKR DLPVTSEDMGRITSCSTNHTPSSDASEWSR GVVVAGQSQAGARVSLGGDGAEAITGLTV DQYGMLYKVAVPPATFSPTGLPSVVMSP LPPTFNVASSLIQHPGIHYPLHYAQLPSTS LQFIGSPYSLPYAVPPNFLPSLLSPSANLAT SHLPHFVPYASLLAEGATPPPQAPSPAHSF NKAPSATSPSGQLPHHSSTQPLDLAPGRMP IYYQMSRLPAGYTLHETPPAGASPVLTPE SQSALEAAAANGGQRPERRNLVRRESEAL DSPNSKGEQGLVPVVECVDGQLFSGSQ TPRVEVAAPAHRGTPDITDLEVQRVVASQV GPQSTILRTQCLCNHLTFFASDFFVVPRTV NVEDTIKFLRVTNPNPVGVSLLASLLGFYV

509

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				ITVVWARKKDQADMQKGCQTPAGVHPPA PQLEEAGTIPSGGLVKVTVLADNDPSAQFH YLIQVYTGYYRSAATTAKLSVYLILPGCRT RTRDPLSGVGSRPVAGAERYLPGQFGRTST VAASNTQAEGAAGHRGFWLAKQHPKDAV TLELRCTPCRSIARLSDAGGVPAGARRVRC AAVLANCSLDMKRGVCASRSATVRKRSD KDVEELGDRESAVGVSDFLDGDHAYERN GNNSHLYQRHKKTKRGVAIARDKMPPDF QDHVIPGQEIKAKSFYSPVDSDETGDKIRY NSKRHRWRTGMLGL
2894	A	3	30	ENFQHFMDRISNGGLEEGKPVLDLVLSCVD NFEARMTINTACNELGQTWMESEVSENAV SGHIQLIIPGESACFACAPLVVAANIDEKT LKREGVCAASLPTTMGVVAGILVQNVLFK LLNFGTVSFYLGYNAMQDFFPTMSMKPNP QCDDRNCRKQEEYKKKVAALPKQEVQIE EEIIHEDNEWGIELVSEVSEELKNFSGPV PDLPEGITVAYTIPKKQEDSVTELTVEDSGE SLEDLMAKMKNM*ISWIE
2895	A	1	2369	AGGARLRPARGRPPRLPPRPGPCRPPVP APTVNERRAPPAGWERRSDAGLSRGARP AEMYGVCGCYGALRPYKRLVDNIFPEDP EDGLVKTNMEKLTIFYALSAPEKLDRIAY LSERLIRDVGRHRYGYVCIAMEALDQLLM ACHCQSINLFVESFLKMOVAKLLESEKPNLQ ILGTNSFVKFANIEEDTPSYHRSYDFFVSRF SEMCHSSHDDLEIKTKIRMSGIKGLQGVVR KTVNDELQANIWDPQHMDKIVPSLLFNLQ HVEEAESRSPSLQAPEKEKESPAELAERC LRELLGRAAFGNKNAIKPVLHLDNHSW EPKVFAIRCFKIIMYSIQPHSHLVIQQLLG HLDANSRSAATVRAGIVEVLSEAAVIAATG SVGPTVLEMFNLTLLRQLRLSIDYALTGSY DGAVSLGTKIIEHEERMFQEAVIKTVGSF ASTLPTYQRSEVILFIMSKVPRPSLHQA VDT GRTGENRNRLTQIMLLKSLQVSTGFQCN NMMSALPSNFLDRLLSTALMEDABIRLVL EILISFIDRHGNRHKFSTISTLSDISVLKLV DKCSRQDTVFMKKHSQQLYRHIYLSCKEE TNVQKHYEALYGLLALISIELANEEVVVDL IRLVLA VQDVAQVNEENLPVYNRCALYAL GAAYLNLISQLTTVPAFCQHIHEVIETRKKE APYMLPEDVFVERPRLSQNL DGVVIELLFR QSKISEVLGGSGYNSDRLCLPYIPQLTDED RLSKRRSIGETISLQVEVESRNSPEKEEVS RATVLGQPHLL
2896	A	1575	1968	REMGFRHVQGOTGLELLTSGDLPTSASQSA GITGVSHHTWPKTLFVLRQSLTSLSPGLECS GTISAHCSPHLPCSSNSCAPASRVAESTEAEH H/LCPDNLHISSREGASPCWPGCS*TPELKR

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				PAHPCRDLGLH
2897	A	524	954	FCSMSSQKWSWQAQPLSWRHWSQGPVPS LPAKLLFKGFLPGTAKPACSAFREAAALAF IQDNKTAISEEKNGSRFLGFPSARLRGRPR AESPRPEPRARPRATQPGPAAPAAHATPPP GPAPAPYLVRGASGGGRGNVRGPK
2898	A	188	590	DLHFEIQVLLBALRGLCSLYPKHREGSLKV HPGHLCWMPVTVRPGTPPSQASTGAQELP GGEKKTCTRWEKKKKTTPGSAGLTGKSIER LTRPALYLRPLLFSSFPVRVTLEALPGGVPK RSASRMPVEMKRGPF
2899	A	41	274	KRGTERKTHFGGCSIQFSDIASGKNILPGLC FLTHKR\WFCSL*RQGWVSRWSHE*GCTR CWRLGKFLWVADRFLGSG
2900	A	1	1462	MKAMPWNWTCLLSHLLMVGMGSSTLLTR QPAPLSQKQRSFVTFRGEPAEGFNHLVVDE RTGHIYLGAVNRIYKLSSDLKVLVTHETGP DEDNPKCYPPRIVQTCNEPLTTNNVNM LLIDYKENRLIACGSLYQICKLLRLEDLFK LGEPYHKKEHYLSGVNESGSVFGVIVSYSN LDDKLFIAATAVDGKPEYFPTISSRKLTKNSE ADGMFAYVFHDEFVASMIPSDTFIIPDF DIYYVYGFSSGNFVYFLTLQPEMVSPGST TKEQVYTSKLVRLCKEDTAFNSYVEVPIGC ERSGVEYRLLQAAVLSKAGAVLGRITLGVH PDDDLLFTVFSKGQKRKMKSLEALCIFI LKQINDRIKERLQSCYRGEGLDLAWLVK KDIPCSSAIRVDGPRGNALQYETVQVVDPG PVLRDMAFSKDHEQLYIMSERQSQELCPPQ ELDDIFSCCQTPRSPDFSHTGTHICALDEAA MAWEWSHSQ
2901	A	14	348	GLFPNKIPFSVLEIRTWAHLSGRHHSACT SCAWPQVACLPLATHPSCTCTFCSLQAPGR PGQSPLSPRRACGPEDLPPPPYV*DLAPSLG PSLGPLMSQSQPRRTPLRG
2902	A	191	1375	EWPEGGGRYSSVPSAVHHARTCLAAELSG TSRPQEPRALPPETGVATAEAEKSNQPAAI SKVNGQGAPLQR/RSPRLSPSPGAAQVPAL PMQDMSEGSSSPGGHIWLASLTPCSLA LWNCCQSPGSQPRGRDEGDCLVRATEPS ATGPDPRRTRLCSSASLVVRNTPDPGISDR RPGISDRRPGTSDRRPGTSDRRPGISDRRPG TSDRRPGTSDRRPGTSDRRPGTSDRRPGISD RRPGTSDRRPGISDRRPGTSDRRPGISDRRPG GTSDRRPGTSDRRPGISRLPRDWIPAAAAS RENSNSADARNRCSSPSRKCTPTSHRMR GSAGSVGSSAGHTAGGTGLPTPSRCSQAL QVFPVAVLGKRGFLSWERSLKQRDIRGPDFS STALI
2903	A	1	2547	MRKYNSLVVDMRKVSVVWIDQASHNIPLS QSQIQIRPFNSVKAERGEEATEEELEANTAS

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GCASRLHSYLLALHCFTVRLCGVPSPHLFA SSTASLPESPGCCMHSLVTKSPCGDPLEPD DATLFKQNLFYLETNLTKQLYHKKIFRTA MLFQFVNVLQVLVHKSHDLLQEEIGIAIY NMASVDFDGFFAAFLPEFLTSCDGV DANQ KSVLGRNFKMDRVCEGISSLSRLQNELSYI EKFTDFLRLFVSVHLRRIESYSQFPVVEFLT LLFKYTFHQDLDIQPSQAVFGGIEFTYILVT LVILGTQRVKPGCGQGGRANCPNSGANA TANGTAAPAAAAAATAYGERPTWRRAD TAGRPATNASASGFPHRIELKAGKTTTLED GRQINGADYLAAPVPGKALAIFGDTGPCD AALDLAKGVDVMVHEATLDTMEAKANS RGHSSTRQAATLAREAGVGKLIITHVSSRY DDKGCQHLLRECRDFKATRPNEKWVTDV TEFAVNGRKLKYLSPVIDLFNNEVISYLSER PVMNMVENMLDQAFKKLNPHHEPVLHSD QGWQYRMRRYQNLKEHGKQSMSRKGN CLDNAVVECFGTLLKSECFYLDEFSNISEL KDAVTEYIEYNSRRISLKLKALAVLANI DPIELTSCADACKRTALVANPWQLGNVR DARTYKELLQIAELLRLGSADRLMEVIR EELELVREQFGDKRREITANSADINLEDLI TQEDVVVTLHQGYVKYQPLSEYEAQRRG GKGKSAARIKEEDFIDRLLVANTHDHILCF SSRGRVYSMKVYQLPEATRARGRPVNL LPLEQDERITAILPVTELGIL
2904	A	165	638	MFVIAFLSPLSLIFLAKFLKKADTRDSRQAC LAASLALALNGVFTNTIKLIVGRPRPDDFY RCFPDGLAHSIDLCTGDKDVVNEGRKSFP SGHSSFAFAGLAFASFYLAGKLHCFTPQGR GKSWRFCAFLSPLLFAAVIALSRTCDYKHH WQGPFW*
2905	A	1	2301	MGWDCGLARWARVGLRERAAVQPLAPG CAAMSFAPFPFIPQGYKTAFGVGTNKIVTQ DNRWELPGA WYFPRASSQAREMPQCPTLE SQEGENSEEKGDSSKEDPKETVALAFVREN PGAQNGLQNAQQGKKRKKRGLGLKAG EWGAMLMIGDQSIQLPAFLSSIVRRAAQQ YGFREGGEDDDWTLYWTDYSVSLERVME MKS YQKINHFFGMSEICRKDLLARNMSRM LKMFPKDFRFFPRTWCLPADWGD LQTYSR SRKNKTYICKPDSCGQKGIFITRTVKEIKP GEDMICQLYISKPFIDGFKFDLRIYVLVTSC DPLRIFVYNEGLARFATTSYSRPCTDNLDDI CMHLTNYSINKHSSNFSRDAHSGSKRKLST FSAYLEDHSYNVEQIWRDIEDVIKTLISAH PIIRHNYHTCFPNHTLNSACFEILGFDILLDH KLKPWLLEVNHSFSTDSRLDKVKDGL LYDTLVLINLESCDKKKVLEERQRGQFLQ QCCSREMRIIEAKGFRAVQLKKTETYEKE

512

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				NCGGFRLLIYPSLNSEKYEKFFQDNNSLFQN TVASRAREEYARQLIQELRLKREKKPFQM KKKVEMQGESAGEQVRKKGMRGWQQKQ QQKDKAATQASKQYIQLTLVSYTPDLLLS VRGERKNETDSSLNQEAPTEEASSVFPKLT SAKPFSSLPDLRNINLSSSKLEPSKPNFSIKE AKSASAVNVFTGTVVSSILEAEKSKIKVLAS LMSGEGFLFDGSFLLCPHTVEGAS
2906	B	1	1518	MVNTERQLDWIERCQVLILALSEEINPELPE AIVMASSEVVTRQDNIDSPQEPPTPLFASR PVTRLKSWRAPRVRPVGPRTHPVVISPVPE CIISIDILRSWQNPHTGLTGRVRAVMVRKA KWKPLELSLPRKIVNQKQYCVPGGIVEISA TTKDLKDAKVVIPTISLFNYPIWLQKNDG SWRMAVDYHKLTTQGVTPIAAAVPNVISLL EQINTSSGTWYAAIYLVNVFFSIPVHKALK KQFAFSWQGPYTFITLPWGHINSPTLCYN LIWRELDHFSLPQDITLVHYIDDIMLIGSSE QEVANTLDLLEKALQQVQAAVQAALPLGP YDPADPVVLEVSADRDTVWSLCSCCYTP WFGTLSHVSNLQWSPCPPVSPVGSQRPO LSREKNKNTKRIHSIPEVLIMKPYFTAVAKP SLLSHKWLPLEKPNPCCYSSDHRTAVPNL LLYRRSTRRKTELTKELTSAHFTGDLPRR AVVVLGDRTAVRPSLEQGMALWI
2907	A	2	266	KGSTAFISGTAGWGTGLLPSSAGLPGGW GPAGGWAGTDRRGPRARPIQKSPWPWS GDAAGQSGFLPVAAWAGQGRLPGGGIIV H
2908	B	494	641	MADLEQLGLNPGLEGTHHLHHPGHMGAK LDKQHPHDRVPTRKSDPACGMGTAVAHH LAPGWLRAAVTQTPFKFCQWKLCSNVIA GDSFSPWYGGISVAHPEPTVTASPTTQGS LPPGEENPSEVVLCAFSKREAEYEHSLRPL KEDRTVYRVGPNKRGKRRTVLKHMOWKL IKGAYRRGQLLANNQAEHKVVSRLINQDC FILEGGTAWKQHALSESSRHALAQFFIVMH LPAQPGALRAPLLLTLAALVHVGVSQSRGS RSRFLGCLEPIERSFLGVLPRSWERSVLCLP VNSLQGAQLRLPAAADSSIFKRS
2909	A	149	300	TRRGGCPEEKVEELKLWEKCVHSLYRHSS SALDLQKIPGAIYIPSGFPLR
2910	B	312	466	MGQVWVLVHSTLEPFHTNNEEEAKYNEV TEEVTEQVCLPAKANAAKEKEVHPYPSAP LNYFEEKEWPPDLSFLEDTGDPSTLTS WQLTKEAEAEQLIEKQVHKAQINRIDPEK IPDLLIFSTQHSPTGVIVQEQLVEWFFLPH TDSWTLTPYLDQITTMIGIGRTRIVKLHGY DPGKIIVPLMKAQIQQAFINSLTWQTHLAD FVGILDNHFPKMKLFQFLKLTNCILPKITKF KPIEGAENVFTDGSSNGKASYFGSKRKVFQ

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				TPY TSAQKVELVA VIELLTAFDMPINVISDS SYVVHSTQLIENAQLRFHTEEKMLTLFTQL QTAVRSRMHRFYTHIRATHLPGSLTEGN QMADRLVATAVSNARHFHSLTHVNASGL KHRYSTWKEAKAIQRCPTCQVVHSSSFT GGVNPRGLEPNSLWEMDVTHVPSFGRLAY VHACVDTFSHFWATCQSGESSAYVKRHL LQCFVVGILASIKTDNAPGYTSQALATFFS IRNIKHITGIPYNSQGAIVERMNLSPETA VAKSKKKGGKQGLRGHPICN
2911	A	3	415	ETGRHRSQQSVSPVPVQPRGKRAMYHSAA ELVSRGFPRPPVQAPAEPAAGAAEGVHSQPA SRQEA/GS/TEVRGQAHFVSPNAAGAGD G/PDPQSLLAPTNRP CPPGGISPARSEPVPPA PGRAAP*CFPDLPGLAPPLC
2912	A	178	423	MLLIPYFLEWKKLWPLAVLSLAWLTYDW NTHSQGGRSAWVRNWTWLVKYFRNYFPV KLVKTHDLSPKHNYIIANHPHGILSF
2913	A	52	228	MLTLPQSLWMLTRTICFVPTIVSCRGLLPS NPHHELARLISVSQHRVWPHVGTQYL*
2914	A	447	1331	SHPLLSCPEKVS AKLRAAAEAAAEERRTR GAGSRGICAGLRVAPGPEPLKQEEGRRE WGSSIGTPSPCGSAQAAAAAAAEATEKIP ALRPALLWALLALWLCCATPAHALQCRD GYEPCVNEGMCVTYHNGTGYCKCEGFL GEYCQHRDPCEKNRCQNGGTCVAQAMLG KATCRCASGFTGEDCQYSTSHPCFVSRPCL NGGTCHMLSRDTYECTCQVGFTGRNPKCP GGNLNYQFNHIVVYSGGSVPPSGTKTSKP AEHNAMGTGSKNFASGLWVMVSGATST STSTL
2915	A	160	409	DSPTSVIWSSTGKYSPHPSAGRVRGYCP RRVLCCPSPEAALEPGRARAQGIRGDSWP HGPTCTQPGRKTVIVGIQLPTQAI
2916	A	1578	685	VFLQQGLAQRITILIGRIYQSWLAIMP GCNH SMTQLHMLSGLRHYHNKSAPVIEVYCPQKP ICKQNWTWLEIMNVFVWEDCIAKQAEVLC NNSYGIIDWSPKGMFSLNCTCQSVCHSHT MFSWSEQNSQMVMVRNTARVPIIWKRG GIVAPQPQMIWSTVEAKHKDLWKLLMSV NKKIWERIKKHLEGHSTNLFLDMAKLKEQ IFKASQAHLTLPMTGVLKGAADKLAASN PLKWMKTLGSSVISMIVLLICVCLCVV CRCRS*LLREVAHRDKAAAFIALQKQEG GYAGE
2917	A	118	399	KWKKYPLGFQTFSSNSQWDTSEFLCSSLL YVLGVSSQNAVNYQSIERSIVGGDCCPFFP WYVHHSWATLKEQRLFLAQQQQEDHEDC TKFEVPH
2918	A	2	335	EDRSAFRPRQPHTLHPLHARSLAPRSPTPPS PPSPDTQLGLSGPTSGPESAPTA/PGNPSWR



Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				SSRWGSSSPCAASST*KSPYP*/CSPT/CAFP SPRLPFCRSAYQPAAGAGRGK
2919	A	486	248	VRQLFSLLLPRLECNGVISAHCNLRPGSC DSSASAS*VARITGASGSQAVVLQVQCLQP VQPGELLRVDLFQLVVLQR
2920	A	3	535	AARQQHCTQVRSRRLMKELQDIARLSDRFI SVELVDESLEFDWNVKLHQVDKDSVLWQD MKETNTEFILLNLTFPDNFPFSPFMRVLSP RLENGYVLDGGAICMELLTPRGWSSAYTV EAVMRQFAASLVKGQGRICRKAGKSKKSF SRKEAEATFKSLVKTHEKYGWGHPARVP DG
2921	A	3384	1260	AGQTPGHRASGPSERSPAPRSRLQPGGEAA TRTEPATPGRAGPGSATMEALMARGAL TGPLRALCCLGCLLSHAAAAPSPIIKFPGDV APKTDKELAVQYLNIFYGCPKESCNLFVL KDTLKKMQKFFGLPQTGDLQNTIETMRK PRCGNPDVANYNFFPRKPKWDKNQITYRII GYTPDLDPETVDDAFARAFQVWSDVTPLR FSRIHDGEADIMINFGRWEGHDGYFPDGD DGLLAHAFAPGTGVGGDSHFDDDELWTL GEGQVVRVKYGNADGEYCKFPFLFNGKE YNSCTDTGRSDGFLWCSTTYNFEKDGKYG FCPHEALFTMGGNAEGQPCKFPFRFQGTSY DSCTTEGRIDGYRWCGTTEDYDRDKKYG FCPETAMSTVGGNSEGAPCVFPFTFLGNKY ESCTSAGRS DGKMWCATTANYDDDRKW GFCPDQGYSLFLVAAHEFGHAMGLEHSQD PGALMAPIYTYTKNFRLSQDDIKGIQELYG ASPDIDLGTGPTPLGPVTPEICKQDIVFDGI AQIRGEIFFFKDRFTWRTVTPRDKPMGPLL VATFWPELPEKIDAVYEAPQEEKAVFFAG NEYWIYSASTLERGYPKPLTSLGLPPDVQR VDAAFNWSKNKKTYIFAGDKFWRYNEVK KKMDPGFPKLIADAWNAIPDNLDVVDLQ GGGHSYFFKGAYYKLENQSLKSVKFGSI KSDWLGC
2922	A	155	575	RRAQGEPEERRAPSLAWTCRDPIPTREELAL TSTTTSCISSLSIVPFQITLVGDSGVGKTSLL VQFDQGKFIPGSFSATVGIGFTNKVGTVDG VREKLPIWTPAGKERFRSVTHAYYRDAHG *FLLYDPNHRJSLRLSAL
2923	C	188	207	MWHLVS
2924	A	3	453	VRSDMNSNPLADGRYRAPPAPRAEAGAS SQP*SPPAAQASGKEGGENNAPLFQ*TPLPT TPTDTLSVPAPRAPVPPSDRFLRSRPPGPRPS FPFRLQGGGGAPH*RGSSATPTPPA/SAPGP GVRSLPRPRWWTPIRLKKPWQKSADPSLQ
2925	A	711	4	GARFACLCSTTPAPMASCLGLLILSSCLLA DCRFIEAWSACTVTCGVGTQVVRIVRCQV LLSFSQSVADLPIDECEGPKPASQRACYAG

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				PCSGEIPFNPDET DGLFGGLQDFDEL YDW EYEGFTKCSESCGGGVQEA VVSCLNKQTR EPAEENLCVTSRRPPQLKSCNLDPCPASSL VVEPKCVGKGHQLFYLT TVLSSRKKQYRL SMERLQRSLLGNQEA WLLILLSPTSSVA
2926	A	2126	2241	RQGFHHVGGAGLKLLTSGDLPALASQSAG IAGMTHSAR
2927	A	830	1143	NDQSALVRARSSFSKSVKPRTHQFFHMFNI GPARDGPPPPSPAPHGPGTLPYRGSSRPGSP PPPPRTTPVSSFLCHSSGAPVTRRDAAAQA HLLCSRFPFSFIG
2928	A	1	782	MTKIQEPSTSVKFLGVQWSGAYQDIPSKV KDKLLHLAPPTTTKEAYLGLFGFWRHQHP H/LGTEQEKTLQHVQAAVQVALFLEPYDP ADPMVLEVSVADRDATWSLWQAPISESQW RPQGFWSKALPSSAANYSPFERQLLAYYW ALVETEHLTMGHQVTKQPELPIMNWVLS PSSHKVGCAQQHSIKWKWYICDRARAGP EGTTTPVITQWAHEQSGHGGRDGGYTWA QQQGLPLTKADLATATAECPICQQRP TLS P
2929	A	1	274	MARATLSAAPSNPRLLRVALLLLLVAAS RRAAGASVVTELRCQLQTLQGIHLKNIQS VNATLKNKGKACLNPA SPMPVQKIEKILN NP
2930	A	1	1236	MLIGSSEQEVANTLDL FVRHLHAREWEIKL TKIQGPSTSVKFLGVQWYGACQDIPSNVK DTLLHLAPPITKKEAQCLLGLFGFWRHQHP HLELPIKNWVLS DPSSYKVGCQQYSIKW KWYICDWAQANPEGTINGLARWSGTWKK HNWKIGDKI WGRGMWMDLSEWSKTVKI YVSHVSAHQMTSAEEDFNNQVDRMTRS MDTTQPLSPTTPVITQWAHEQSDHGGRDG DYTWAQQHGLPLTKSFTFAKEVWQWAHA HGIHWSYVPHHPEAAGLIERWNGLLKSQL KCQLGDNTLQGWGKVLQKAMYALNQHP YGTVSP IARLHGSRNQGEVEVAPLIITPGD LLAKFLLPVSTTLHSAGLG VVYGFKLTRD GLVMVNT ECQLDRIEGCKVFLGVSVRVS PKEINI
2931	A	3	714	RRPFIALCLSNVAFMLPWQFAQFILFTQIAS LFPMYVVGYIEPSKFQKIYMN MISVTL SFI LMFGNSMYLSSYYSSLLMTWAILKRNEI QKLGVSKLNCWLIQGS AWWCGTHLKFLT KILGVSDHICLSDLIAAGILRYTDFDTLKYT CSPEFDFMEKATLLIYTKTLLPVVMVITCF IFKKT VGDISRVLATNVYL RKQLLEHSELA FHTLQLLAFTALAILRLKLV
2932	A	1	699	MRFVMSVTMYHTTLVGLDIKHLNLESGKV WVMGKASKEPRLPIGRNAVAVIEHWLIDL RDLFGSKDDALFLSKLGKRISARNVQKRFA

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				EWGIKQGLNNHVHPHKL R H S FATHM LESS GDLRGLFRFVSAKRHAKGSKVGSPIIYADQ IIGAGQNH PARWRGLPRKSRL LVSPSNDK RRKAGAAPVAALRHFPPISENAIVKIQFRRI RRLNHQQLVKPYPQVPISQATNQFR
2933	A	1	924	MFAIISYSSLA AVL TATLTAAGIISFPVALC LVIGANLGSGLLAMLNNSAANAARRVAL GSLLFKLVGSLIILPFVHLLAETMGKLSLPK AELVIYFHVFNLVRLVMLPFVDPMARF CKTHRDEPELDTQLRPKHL DV SALDTPTLA LANAARETCALATPWTDDGRKYA YSAAS GGRRSATKVMVVVTDGESHDGSM LKAVI DQCNHDNLRFGIAVLGYLNRNALDTKNLI KEIKAIASIPTEYFFNVSD EAALEKAGTL GEQIFSIEDMDLGDEVYTVGRPHPMIDPTL RNQLIADLGAKPQVRVLLLDVVIGFGATA DPAASLVSAWQKACAARLDNQPLYAIATV TGTERDPQCRSQIATLEDAGIAVSSSLPE ATLLAAALIHPLSPAAQHTPSLLENVAVI NIGLRSFALELQSASKPVVHYQWSPVAGQ GKWLANPELLEADADA EYAAVIDIDLADI KEPILCAPNDPDDARPLSAVQGEKIDEVFIG SCMTNIGHFRAAGKLLDAHKGQLPTRLWV APPTRMDAAQLTEEGYYSVFGKSGARVSSI PCAVPCVWARVADGATVVSTSTRNFPNRL GTGANVFLASAEAAVAALIGKLPTPEEYQ TYVAQVDKTAVD TYRYLNFNQLSQYTEK ADGLLKPRFRPWQRKILDTLATYHEQHRD EPGPGRERLRRMALPMEDEALVLLIEKM RESGDIHSHHGWLHLPDHKAG*SSDNKY QRLFYLPAPRRSGTLPASAVCQSAPQQ/LA SSAEARKTFAPVPRRFGKLRVEVETTVAPS ATRAHTQGTAQGILDTRAPLLPKTL
2934	A	201	632	MPGLLNWITGAALPLTASDVTSCVSGYAL GLTASLTYGNLEAQPFGFLVYPLDECTTV IGFEAVIADRVVTVQIKDKAKLES GHFDAS HVRSPVTGNILQDGVSIAPHSCTPGKVTL DEDLERILFVANLWTIAPMYRAVWD
2935	A	267	25	MGA VQRLMKIIMLNRYLVAHFLVLF A QK KANRQRTRVHRGSLWLSECESPNPGGRH TEPAEGRQARGRT PQQGF AVSLM*
2936	A	34	330	MNKHFLFLFLLYCLIAAVTSLQCITCHLRT RTDRCRRGFGVCTAQKGEACMLLR IYQRN TLQISYMCQKFCRDMTFDLRNRTYVHTC CNYNYCNFKL*
2937	A	34	411	MTAGTVVITGGILATVILLCHIAVL CYCRLQ YYCCKKSGTEVADEEEEREHDLPTHPRGP TCNACSSQALDGRGSLAPLTSEPCSQPCGV AASHCTTCSPYSSPFYIRTADMVPNGGGGE RLSFAP
2938	A	333	545	MMPTNLAHLVFWQALLASGRFSLMEHYP

517

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				PNVQSNRGITHYMLPRGYILGLLYSSAGNT GTSRPRRTHYGT*
2939	A	242	382	MNRVMRGLAITTTCLLSMLQAITSPILW NHAAVQYVHGHSLVQA*
2940	A	108	290	MPQWLALQRQQALLTLLSGAGTWAGMRP PSQCWPQGPSTGNQSLSHGRGELLTHAVG VCI*
2941	A	109	417	MLMLILVTGVSSLRNMIMCDYISRAKLKSS HIVLSYCTLKQBYDDSRGVMNLEAREEGS RGFYCLGCIDTGLQTPGGRGPSSALVTSVH LACEEYSKHSFVK*
2942	A	155	575	RRAQGEPEERRAPSLAWTCRDPIPTREELAL TSTTTSCISSLSIVPFQTLVGDSGVGKTSLL VQFDQKGKFGSFSATVGIGFTNKVGTVDG VREKLPIWTPAGKERFRSVTHAYYRDAHG *FLLYDPNHRISLLRLSAL
2943	A	429	1	RLVYASTANKIHF*NDNNPGKNTDTPVPHC HKLCNQDSHIRGNHRGQHIHSKTAKPCSG KTTFVIITFLLSDKHKYKLAPLRPAAASYSS PFRKVTCLTRITEPS*P*HTAATLRSDQRS QTCSHGTGTLSWRSSRWSSSTK
2944	A	1728	2782	RASSAVRGSGLGDSARGRRRRSIVKVSHPA VMSKSESPKEPEQLRKLFIGGLSFETTDESL RSHFEQWGILTDCVVMRDPNTRSRGFGF VTYATVEEVDAAMNARPHKVDGRVVEPK RAVSREDSQRPGAHLTVKKIFVGGIKEDTE EHHLRDYFEQYQKIEVIEIMTDRSGGKKRG FAFVTFDDHDSVDKIVIQKYHTVNGHNCE VRKALSKQEMASASSSQRGRSGSGNFGGG RGGGFGGNDNFGRGNGFSGRGGFGGSRG GGGYGGSGDGYNGFGNDGSNFGGGGSYN DFGNYNQSSNFGPMKGGNFGGRSSGPYG GGGQYFAKPRNQGGYGGSSSSSYGSGRR F
2945	A	234	657	VQQPGRGLDLSTDGPGRSQVGLIWSCCC LH*AASGEPGGRCPGS/GAPGPAGSALEFR ARDGVP\GVGGPSWESHSPAAATPPPAECR GPGPTPSPAPGEAAPEDREDGAAAPGRAEP ASIVAPADGSQGGVLAATQAGALGA
2946	A	1725	2140	YTYQISQTSGL*PGDKSVHSELV/SSCNTSI ISSSGISSTSL*LRRLFSAASANSASSVASK K*ASSMPLSQASADAPVDSLLGDGL*GF WVSLLLVSSASSVNNSSSLKKNRRHTSAG NGKQSDLKFFALHTGS
2947	A	1	1134	DTYCRGDQLHILLVVRDHLGRRKQYGGDF LRARRSPALMAGASGKVTDNFNGTYLVS FTLFWEGQVSLSLLIHPSEGVSALWSARN QGYDRVIFTGQFVNGTSQVHSECGLILNTN AELCQYLDNRDQESFYWVRPQHMRCAAL THMYSKNKKVSYLSKQEKSLFERSNVGVE IMEKFNTISVSKCNTLKSVDLHESGKLQHQ

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LAVDLDRNINIQQWQKYCYPLIGSMTYSVK EMEYLTRAIDRTGGEKNTVIVISLGQHFRP FPIDVFIRRALNVHKAIQHLLRSPDTMVII KTENIREMYNDAERFSDFHGYIYLIKDIF QDLSDIRHVLKYNASKNAADLDLFFSSNL DDFYNFSELHKGRSKSPLMQITQ
2948	A	504	198	QLIQHQT VHTGRKLYECKECGKAFNQGST LIRHQRIHTGEKPYECKVCGKAFRVSSQLK QHQRHTGERPYQCKELKGRGAEMLA VLA VKEQNRTPVNYGK
2949	A	1	578	MGETALMIQLPPPGPALGTWGLWDLQFKT NTTSTDTPRSHLQETGDNILTLFTMHPPPL ESEWTICNFRQIWLSSWSTLETRAQPLHS YFRKLKGRGTALAGIVFGIVFIMGVIAIAI CICMCMKNHRATR VGILRTTHINTVSSYPG PPPYGHDHEMEYCADLPPPYSPTPQGAQR SPPPYPGNARK
2950	A	1	943	AAAGRARGAGDMFRRKQSNPRQIKRSLGD MEAREEVQLVGASHMEQKATAPEAPSPPS ADVNSPPPLPSPTSPGGPKELEGQEPBRPT EBEPGSPWSPDELEPVVQ/DGRRRRIRARLS LATGLSWGPFHGSVQTRASSPRQAESPAL TLLLVDEACWLRTLQALTEAEANTEIHRK DDALWCRVTKPVPAGLLSVLLTGEPHST PGHPVKKEPAEPTCPAPAHDLQLLPQQAG MASILATAVINKDVFPCKDCGIWYRSENL QAHLLYYCASRQGTGSPAAAATDEKPKET YPNERVCPYPQSRKSCPG
2951	A	2	435	AVCRTSSDVDDNPPVFNQLIYESYVSELAP RGHFVTCVQASDADSSDFRLEYSILSGND RTSFLMDSKSGVITLSNHRKQRMEPLYSLN VSVSDGLFTSTAQVHIRVLGANLYSPAFSQ STYVAEVRENVAAGTKVIHVRATD
2952	A	199	399	MPGSLCGRRTVCWLLGSVTSKQVLTDFLR KFSRSSRLQEDQERSLGFRPFTHSPDMMW DLPAQDEWS
2953	A	38	397	TVLCLTLTSCSFRQSLAT*SFGG/MGSGSVH FGVGGAFLPSIHWS/GSRSLSVSSTHFVP SSSS/GGYGSGDASVLCRSDRLLTGTKITTQ NIHD/RLGSYLDKVRAL EEAGELKV KICD WAP
2954	A	2	673	NSRVEGQLCDLDPSAHFYGHCGEQLECR DTGGDLRGEVPEPLCACRSQSPLCGSDGH TYSQICRLQEAARARPDANLTV AHPGPES GPQIVSHPYDTWNV TGQDVIFGCEVFAYP MASIEWRKDGLDIQLPGDDPHISVQFRGGP QRFEVTGWLQIQAVRPSDEGT YRCLARNA LGQVEAPASLTVLTPDQLNSTGIPQLRSLN LVP EEEAESEENDYY
2955	A	1	440	GNQKCTRNNHRISSLLCDPQEGYLQMLQIS NLYLYDSVLMLANAFHRKLED RKWHNM

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				ASLNCIRKSTKPWNGGRSMLDTIKKGHTG LTGVMEFREDSSNPYVQFEILGTTYSETLAE EPFVMVAENILGQPKRYKGFSDVLDALA
2956	A	23	395	GSGDAGGQHRARCPSGRAGNWDWHPPA MEEPGPPGGLSQDQVERCMGAMQEGMQ MVKLRGGSKGLVRFYYLDEHRSCIRWRPS RKNEKAEISIDSIQEVSEGRQSEVLQRYPDG SFDPNCCCSI
2957	A	663	144	KELSAVSAGIPHSCGSQCGGGSVAAACVP AAPAAAGLCSGRAQKVPPPPSLAGWPPGV NAPPPVPCSSVRLHVCQSDRLWVRLAARR GILALLRSALKAAATLAGCQSVRWSVRPSES LRPTSNAASLFRSSVPTVLSHSVPLAASLG KRRACGGREHASVAVYLSVCLSLPT
2958	A	1856	591	PPTPTAETLTSEDAQPGSPLATGTDQVSLD KPLSSAAHLDDAAKMPSASSGEEADAGSL LPTTNELSQLAGADSLDSPRPLERSVGQ LPSPPLPTPPPKASSKTKKMSQAKPHSSK PPA*RVPTL/PLRGQLSTPTGSPHLTTVHRP LPPSRVIEELHREALTKHRQDSFQGRESKG SPKKRLDVRLSRTSSVERGKEREEAWSFD GALENKRTAAKESEENKENLIINSELKDDL LLYQDEEALNDSIISGTLPRKCKKELLAVK LRNRPSKQELEDNRNIFPRRTDEERQEIRQI EMKLSKRLSQRPAVEELERRNILKQRNDQ TEQEEERREIKQRLTRKLNQRPTVDELDRK ILIRFSDYVEVAKAQDYDRRADKPWTRLS AADKAAIRKELNEYKSNEMEVHASSKHLT RFHRP
2959	A	1578	685	VFLQQGLAQRITILIGRIYQSWLAIMPGCNH SMTQLHMLSGRLIYHNKSAPVIEVYCPQKP ICKQNWTWLEIMNVFVWEDCIAKQAEVLC NNSYGIIDWSPKGMFSLNCTCQSVCHSHT MFSWSEQNSQMVMVRNTARVPIWKRG GIVAPQPQMIWSTVEAKHKDLWKLLMSV NKIKIWERIKKHLEGHSTNLFDMAKLKEQ IFKASQAHLTLMPGTGVLKGAADKLAASN PLKWMKTLGSSVISMMIVLLICVVCLCVV CRCRS*LLREVAHRDKAAFAFIALQKQEG GYAGE
2960	A	470	258	MIIAGGVIVASGLVFIVLLMIRYKVYGDG DSRRVKGSRALPRVRHVCSQTNGAGTGAE QAPALPAQDHY*
2961	A	3	866	ELNLQDFSHLDHRDLIPILAALEYNQWFTK LSSKDLKLSTDVCEQILRVVSRNREELV LENAGLRDFAQKLASALAHNPNSGLHTI NLAGNPLEDRGVSSLSIQFAKLPKGLKHLI LSKTHYYPKAVNSLSQSLSANPLTASTLVH LDLSGNVLRGDDLSHMYNFLAQPNIVHL DLSNTECSLDMVWGALLRGCLQYLAVLN LSRTVFSHRKGKEVPPSFKQFFSSSLALMHI

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				NLSGTKLSPEPLKALLGLACNHNKGVSL DLSNCELRSQGAQVLEGCIG
2962	A	574	203	TQAFEQEVGNPLCIPSHCMGAVFILLNLAT AHSSGLCLLQLELSFRSLSTAVHCCPRPTI DFHP/LGSSRVSAVLLIQ/QRCPLPLIGLEA DHCSMAKGPFGFILIELNTSHWVPQFSSVT HDFY
2963	A	399	15	NTMVAHHIVENTYFCPVLATGLSGLYSSLP TKLEEKGEWHCLLKDDWLLPSLVQFM NSLEFCNAVIQVAHPLIRNQLVIYISNEFLV PVLAPALHKVPVQEVMSPTAYLDFVRSIS EPALLEIF
2964	A	3	567	CSEIFASRLPRIMAHSKQPSHFQSLMLLQ WPLSYLAIFWILQPLFVYLLFTSLWPLPVL YFAWLFLDWKTPERGGRSAWVRNWCV WTHIRDYFPITILKTKDLSPEHNYLMGVHP HGLLTFGAFCNFCTEATGFSKTFPGITPHLA TLSWFFKIPFVREYLMAGASDHTYWSFW SMFLLGNAPF
2965	A	2	394	TLADGGEGQFDGTFEPATVALPGGEHAEN AVQIHKVVTGTMALIFSFLIAALVLYVSWK CFPASLRQLRQCFVTQRRKQKQKQTMHQ MAAMSAQEYYVDYKPNHIEGALVIINEYG SCTCHQOPARECEV
2966	A	2	412	EFLSSNQITQLPNTTFRPMPNLRSDLSYN KLQALAPDLFHGLRKLTTLHMRANAIQFV PVRIFQDCRSLKFLDIGYNQLKSLARNSFA GLFKLTLEHLEHNDLVKVNFAHFPRILSLH SLCLRRNKVAIVVSSLDW
2967	A	1	1343	ERCKVQSSTLVSSLEAELSEVKIQTHIVQQE NHLLKDELEKMKQLHRCPLSDFFQKISS VLSYNEKLLKEKEALSEELNSCVDKLAKSS LLEHRIATMKQEKSWEHQASLKSQLV SQEKVQNLEDTVQNVNLQMSRMKSDLRV TQQEKBALKQEVMSLHKQLQNAAGGKSWA PEIATHPSGLHNQKRLSWDKLDHLM/NV EEQQLWQENERLQTMVQNTKAELTHSRE KVRQLESNLLPKHOKHLNPSGTMNPTEQE KLSLKRECDQFQKEQSPANRKVSQMNSLE QELETIHLENEGLKKKQVKLDEQLMEMQH LRSTATPSPSPHAWDLQLLQQACPMVPR EQFLQLQRQLLQAERINQHLQEELENRTSE TNPQGNQEQLVTVMEEERMIEVEQKLKLV KRLLQEKVNQLKEQVSLPGHLCSPSSHSSF NSSFTSLYCH
2968	A	382	203	RPSSPGPPCPEAGKR/RFGCGGAGSLRPEHS VTRPPRGLGKGRGQREKRGASKEGSEGCA
2969	A	303	46	AVVFKLLSPRKKHLKNPFVGGVGCAWRT GWEWSPGQEQAPPPATGSMLATSSPPSGPP PPP*PPGFMLPPLGDGLGAGTSAGRS*EKG RGK

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2970	A	3	586	MVECPACQH*RPTLSLRDTSYHQVECSRSL LPWNGHQFVLTRIDICK*/G/FVFPNYFASS STTI*ELTGCLIH*HT*N*GTH/LIAKEV*Q*T RSYKI/HWCI/PHHPEAASQIGFWNGLLKTG L/QLRLRCNALQS/WGAVLQNMVYALKCI GPKWIYSIVSPVGHVHTGVASTITTPSHSPV EFVPPRSEIWSQLGYDP
2971	A	299	21	MGSSVLSIWILSPSIYPILSPLAMPCLSRDIL IRVRRIQGAWPSEGTAASSIRGWVLTCLRMS SGKALEALYCIPGAAQHPGLGVTRVWSGR T*
2972	A	1	555	KKVGNYYTTPYRFRMKCHLCVNYIEMQT DPANCDYVIVSGAQRKEERWDMADNEQV LTTGERHPLTCLGAL/DPESALGPPKPSRAL IVAEHEKKQKLETDAMFRLEHGEADRSTL KKALPTLSHIQEAQSAWKDDFALNSMLRR RFRVRGAPARGQRGCMVDQGPALPPPH PSFEQATCTF
2973	A	1	598	MAVVIPAALGTAALVPWSILRGKAPRYWL LPLLLDPDKVPIISARDLTSPDAALASLTAQ SGGLEELHLKLVHEVAVMANTECOLDWIE GCKVLILACRLWDLVIMTHPAFYQSVQWG KGNDQTFQGRDLTGCELMIPGDPNCGPP VKVG VYGGHIYHCDLTKEELEPRVFREVT KGIDASDYQTVQLPKGTESSRN
2974	B	1	2142	MGGAGSPQVILVSHTPQSASAACEEIAYQV AGVSGNLAPGNQPEKEGRAHQCLECDRAF SSAAVLMHHSKEVHGRERIHGCPVCRKAF KRATHLKEHMQTHQAGPSLSSQKPRVFKC DTCEKAFAPKPSQLERHSRIHTGERPFHCTL CEKAFNQKSALQVHMKKHTGERPYKCAY CVMGFTQKSNMKLHMKRAHSYAVAVAM GGTAQCPPGATACLGTAICPSGLRAQPSN LSVPEAAKPKSGRNRKIEAPTWALSTSKDP QTEGLRNPQTCVQIRSNPFCFAAQGFLISE LRTLNCVFVGLCDSQSGKQQLGFYSGPAT EAWQKYS LAVCILRSEQEISATRLGLKNTN VNKLDGGCGAWNFLGGMSEHNSPPSGRAI LLPVVFTEVFPWPWTPEQGSICRMNLAPT FQAFLPKTGFPIDPQELLQGPIERTWPGTV YTFRSAIVTARAVWVRPRMDRRADLSSAT QSASAEKFGGRVSAGHCALPLPARPVTAS VYGRLARLRGCLEDSYPSALSAQVFLDSPA VGCGLETRLFIEAALGPPCRATVTSRGHLL DISITKSPGRPCFLSVCLHGSQQKRGAA ATAKRKSKGGGVNVEGRLCTWPPEDPPKS WSLAFGPLQEKTTTELNLHPRCWARCLSHW ELPPGPRGRAQAPDWTGSKSFREQLLFTL WGVQEKISKHQANQGKEAPAYTGLEDSDP GGLCAV*
2975	A	248	597	DRCPAAWDRHPAGIQSSRREPSKATWTLR



Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				SKLSVQDGRDSSLRLNCKVAARLGAGHP PMLRLGLRC*YPGKQGLEWTSSKLQQTCH *GS*LLKGKLTNRKDIHDKTPSVRHYHQR
2976	A	2	353	EVDHRGDYVSHEIMHHQRRRRRAVPVSEVE PLHLRLKGSRHDFHVDLRTSSSLVAPGFIV QTLGKTGKTSVQTLPPEDFCFYQGSLRSHR NSSVALSTCQGLSGMIRTEEADYFLRPL
2977	A	134	412	MVKFIGPRVRRGLESPLCHACYLALCTLAL VRLCALSRSRSLMLILQAFYRPPMSQEP ALSTVLFLLLLLANPPTKVSRSRHKERVLL LVA
2978	A	1	598	MAFLETSAPLYEHIWTLQVAFSTVGLGETL KVAMISMSTSSGYFLQLQYCCSSTITITGY KGFLRDLKVETRADGVMRTMAPEKLLKS MPILOQQIDALLEFDVHPNELTNGVINAAF MLLFKDLIKLFACYNDGVINLLGTWMKLE THLSKLLQRQKTKHCMFSLIGNRTMRTL GHRKGNITHWALLAGGGAAEG
2979	A	793	1	GSRIDDMKSERRPPSPDVTVLSDNEQPSSPR VNGLTTVALKETSTEALMKSSPEERERMIK QLKEELRLEEAKLVLLKKLRQSQIQKEATA QKPTGSGVSTVTPPPLVRGTQNPAGKPS LQTSSARMPGSVIPPLVRGGQQASSKLGP QASSQVVMPPPLVRGAAQQIHSIRQHSSTGPP PLLAPRASVPSVQIQGQRHQQGLIRVANV PNTSLLVNIPQTPASLKGTTATSAQANSTP TSVASVVTSTESPASRQAA
2980	A	2	1427	LLARGAGRTNPAPPLMSCGPWGKFLKCCE VYKSGPYKVQ*EEITHSRABAESTYQIKYE ELQTLAGKHGDDLRCAT/EISEMNQNISR LQAETEGKGGASLEAAIADAEQWGELA IKDANTKLSELEAAMQRAKQDMA/RQLGE YQKLALDIELATYRKLLGEESRLESQM VSIHKTTSGYAGAPARIVSLLQNELLSLE VGVLKGHPTGKGEELGAPYSECSFGLCRR TVMLTQAPSSVVRSRNHTVNSGGSC SASTVAIPAINSSAAMSACSTISAQKRTCC TACEPARKYKDTASHQEPAVCQPACQLET ADPKGGGVLPALPQPPSPGMLCWPYCR ATDYFLANFFSEFPCHFLHRAGAAQTQAT GDGMEHGQSRELPRKAPREESETSEEKSP NKWGPVSKQKKQLLDILTTHRPTRGNAY TGLSTRKWKPRSEENALMQPNKKDEKGT L
2981	A	4235	940	ARGRRSRPVWAASWGGRRPAARRRPRG LAATMGFELDRFDGVDPLKCALCHKV LEDPLTPCGHVFCAGCVLPWVQEGSCP ARCRGRLSAKELNHVLPKRLILKLDIKCA YATRGCGRVVKLQQLPEHLERCDFAPARC RHAGCGVLLRRDVEAHMRDACCARDPVG RCQEGCGLPLTHGEQRAGGHCCARALRA

523

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				HNGALQARLGALHKALKKEALRAGKREK SLVAQLAAAQLELQMTALRYQKKFTEYSA RLDSLSRCVAAPPGGKGEEKSLTLVLHRD SGSLGFNIIGRPSVDNHDGSSSEGIFVSKIV DSGPAAKEGGLQIHDRIEVNDRDLRATH DQAVEAFKTAKEPIVVQVLRRTPRTKMFT PPESQLVDTGTQTDITFEHIMALKMSSPS PPVLDPYLLPEEHPSAHEYYPNDYIGDIH QEMDREELELEEVDLYRMNSQDKLGLTV YRTDDEDDIGIYISEIDPNSIAAKDGRIREG DRIQINGIEVQNREEAVALLTSEENKNFSL LIARAEQLDEGWMDDDRNDLDDLHMD MLEEQHHQAMQFTASVLQKKKHDEGGT TDTATILSNQHEKDSGVGRDSTRNDESS EQENNGDDATASSNPLAGQRKLTCSDTL GSGDLPFSNKSFSPECTGAAAYLGIPVDECE RFRELLELKQCVKSATPYGLYYPSPGLDAG KSDPESVDKELELLNEELRSIELECLSVRA HKMQLKEQYRESWMLHNSGFRNYNTSI DVRREHLSITELPEKSDKDSSAYNTGES CRSTPLTLEISPDNSLRRAAEGISCPSSGA VGTTEAYGPASKNLLSITEDPEVGTPTYS LKELDPNQPLESKERRASDGSRSPTPSQKL GSAYLPSYHHSYPYKHAHIPAHAQHYQSYM QLIQKSAVEYAQSQMSLVSMCKDLSSPT PSEPRMEWKVKIRSDGTRYITKRPVRDRL RERALKIREERSGMTTDDDAVSEMCMGR YWSKEERKQHLVKAKEQRRRREFMMQSR LDCLKEQQAADDRKEMNILELSHKKMMK KRNNKIFDNWMTIQELLTHGKTPDGTRV YNSFLSVTTV
2982	A	792	389	PTRPPLQLQAPRAHLSAQKRLMLMKQKG VMNQPMAYAAPLPSHGQEQHPVGLPRTTG PMQSSVPPGSGGMVSGASAPGFLGSQP QAAIMKQMLIDQRAQLIEQQKQFLREQR QQQQQQQILAEQVTCPLA
2983	A	3	268	FTRSDELAHYRTHHTGEKRFSCPLCPKQFS RSDHLTKHARRHPTYPDMIEYRGRRTTP RIDPPLTSEVESSASGSGPGEAPSFTTCL
2984	A	3	431	GPEFPGSAKLVLFDLSYNNLTQLGAGAFRS AGRLVKLSLANNLGVGHEDAFETLESQ VLELNDNNLRSLVAALAALPALRSLRLD GNPWLCDCDFAHLFSWIQENASKLPKGLD EIQCCLPMESRRISLRACRRPASRV
2985	A	108	497	MGIYQMYLCFLAVLLQLYVATEAILIALV GATPSYHWDLAELLPNQSHGNQSGAGEDQ AFGDWLLTANGSEIHKHVHFSSSFTSIASE WFLIANRSYKVSAASSFFSGVFVGVISFG QLSDRFGRKKVY
2986	A	488	754	QSIYQEKFDDENFILKHTGPGILSMANAGP TQMVPSPVWPRLSGWMASTRSLAK*EE

524

Table 8

SEQ ID NO:	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GVNIMEAMECSGSGNGETGKKIPTAXCGQ L

525

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
1	1042			
2	1043			
3	1044			
4	1045			
5	1046	2083	2535	790 104
6	1047			
7	1048			
8	1049			
9	1050	2084	2536	790 16362
10	1051			
11	1052			
12	1053			
13	1054			
14	1055			
15	1056			
16	1057			
17	1058	2085	2537	784 5743
18	1059	2086	2538	790 167
19	1060			
20	1061	2087	2539	788 2001
21	1062			
22	1063	2088	2540	784 1683
23	1064	2089	2541	785 1699
24	1065			
25	1066			
26	1067	2090	2542	789 5434
27	1068			
28	1069	2091	2543	790 13996
29	1070			
30	1071			
31	1072			
32	1073			
33	1074	2092	2544	784 6213
34	1075	2093	2545	784 1993
35	1076			
36	1077	2094	2546	790 3341
37	1078	2095	2547	791 5740
38	1079			
39	1080	2096	2548	792 4643
40	1081			
41	1082			
42	1083			
43	1084	2097	2549	790 407
44	1085			
45	1086	2098	2550	785 1457
46	1087	2099	2551	790 20129
47	1088			
48	1089	2100	2552	790 18963
49	1090	2101	2553	790 515
50	1091	2102	2554	787 7703

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
51	1092			
52	1093			
53	1094	2103	2555	784 7239
54	1095	2104	2556	790 19031
55	1096	2105	2557	791 1750
56	1097			
57	1098			
58	1099			
59	1100	2106	2558	790 23024
60	1101			
61	1102	2107	2559	788 3666
62	1103			
63	1104	2108	2560	787 2031
64	1105			
65	1106			
66	1107	2109	2561	784 2939
67	1108	2110	2562	787 4769
68	1109	2111	2563	792 7097
69	1110	2112	2564	788 9897
70	1111	2113	2565	790 29652
71	1112			
72	1113	2114	2566	784 4530
73	1114			
74	1115			
75	1116	2115	2567	787 7560
76	1117			
77	1118			
78	1119			
79	1120			
80	1121			
81	1122			
82	1123			
83	1124	2116	2568	784 1264
84	1125	2117	2569	791 1515
85	1126			
86	1127	2118	2570	784 3498
87	1128			
88	1129			
89	1130			
90	1131			
91	1132			
92	1133			
93	1134	2119	2571	791 1404
94	1135			
95	1136	2120	2572	784 9584
96	1137			
97	1138	2121	2573	787 7852
98	1139			
99	1140	2122	2574	788 5026
100	1141			
101	1142	2123	2575	790 16594

527

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
102	1143	2124	2576	790 975
103	1144			
104	1145			
105	1146			
106	1147			
107	1148	2125	2577	790 11619
108	1149	2126	2578	790 1040
109	1150	2127	2579	787 946
110	1151			
111	1152			
112	1153			
113	1154	2128	2580	790 19602
114	1155			
115	1156	2129	2581	788 12191
116	1157	2130	2582	784 5727
117	1158			
118	1159	2131	2583	784 7669
119	1160			
120	1161	2132	2584	784 5053
121	1162			
122	1163			
123	1164			
124	1165	2133	2585	790 9619
125	1166			
126	1167			
127	1168	2134	2586	790 1144
128	1169			
129	1170			
130	1171			
131	1172	2135	2587	790 16699
132	1173	2136	2588	790 1170
133	1174			
134	1175	2137	2589	790 1171
135	1176			
136	1177			
137	1178			
138	1179			
139	1180	2138	2590	785 66
140	1181	2139	2591	790 11744
141	1182			
142	1183			
143	1184	2140	2592	784 10222
144	1185	2141	2593	790 1217
145	1186	2142	2594	785 2455
146	1187			
147	1188			
148	1189	2143	2595	784 3575
149	1190			
150	1191			
151	1192			
152	1193	2144	2596	787 9817

528

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
153	1194			
154	1195	2145	2597	784 9353
155	1196			
156	1197			
157	1198			
158	1199	2146	2598	784 4306
159	1200			
160	1201			
161	1202			
162	1203			
163	1204	2147	2599	790 23831
164	1205			
165	1206			
166	1207			
167	1208	2148	2600	790 1363
168	1209	2149	2601	784 1344
169	1210			
170	1211			
171	1212	2150	2602	787 1542
172	1213			
173	1214	2151	2603	785 2871
174	1215	2152	2604	787 5391
175	1216	2153	2605	790 27456
176	1217			
177	1218	2154	2606	784 1229
178	1219			
179	1220	2155	2607	788 1187
180	1221	2156	2608	784 256
181	1222			
182	1223			
183	1224	2157	2609	790 6023
184	1225			
185	1226	2158	2610	790 28512
186	1227			
187	1228			
188	1229			
189	1230			
190	1231			
191	1232			
192	1233	2159	2611	790 27560
193	1234	2160	2612	784 9678
194	1235			
195	1236	2161	2613	787 2238
196	1237			
197	1238	2162	2614	787 8011
198	1239			
199	1240	2163	2615	784 9436
200	1241	2164	2616	787 6897
201	1242			
202	1243			
203	1244	2165	2617	790 1649

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
204	1245			
205	1246	2166	2618	790_1664
206	1247	2167	2619	790_1671
207	1248	2168	2620	789_4182
208	1249	2169	2621	787_3365
209	1250	2170	2622	790_24699
210	1251			
211	1252	2171	2623	790_24002
212	1253			
213	1254	2172	2624	790_1713
214	1255			
215	1256	2173	2625	790_12005
216	1257			
217	1258	2174	2626	787_371
218	1259	2175	2627	788_11375
219	1260	2176	2628	792_6253
220	1261	2177	2629	790_20480
221	1262			
222	1263	2178	2630	787_8084
223	1264			
224	1265	2179	2631	790_1787
225	1266	2180	2632	787_5659
226	1267	2181	2633	790_14480
227	1268	2182	2634	790_1801
228	1269			
229	1270	2183	2635	790_22521
230	1271	2184	2636	790_3633
231	1272			
232	1273	2185	2637	787_5670
233	1274	2186	2638	790_20482
234	1275			
235	1276	2187	2639	790_6685
236	1277	2188	2640	785_2624
237	1278			
238	1279			
239	1280	2189	2641	787_6797
240	1281	2190	2642	784_5046
241	1282			
242	1283			
243	1284			
244	1285			
245	1286			
246	1287			
247	1288	2191	2643	784_6709
248	1289			
249	1290			
250	1291	2192	2644	787_3930
251	1292			
252	1293	2193	2645	790_2982
253	1294	2194	2646	790_2086
254	1295			



530

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
255	1296			
256	1297			
257	1298			
258	1299	2195	2647	784 1280
259	1300			
260	1301	2196	2648	787 9953
261	1302	2197	2649	790 4258
262	1303	2198	2650	790 16925
263	1304	2199	2651	790 1256
264	1305	2200	2652	788 6514
265	1306			
266	1307			
267	1308			
268	1309			
269	1310			
270	1311			
271	1312			
272	1313	2201	2653	787 2484
273	1314	2202	2654	790 2283
274	1315			
275	1316	2203	2655	787 2505
276	1317	2204	2656	790 6292
277	1318			
278	1319			
279	1320	2205	2657	784 2332
280	1321			
281	1322			
282	1323	2206	2658	790 2410
283	1324	2207	2659	790 6347
284	1325	2208	2660	790 12379
285	1326	2209	2661	790 2433
286	1327	2210	2662	784 8177
287	1328	2211	2663	790 2436
288	1329			
289	1330			
290	1331			
291	1332	2212	2664	790 2469
292	1333	2213	2665	788 7
293	1334	2214	2666	784 6493
294	1335			
295	1336			
296	1337	2215	2667	790 2489
297	1338			
298	1339			
299	1340	2216	2668	790 8006
300	1341	2217	2669	787 2576
301	1342	2218	2670	790 2537
302	1343			
303	1344	2219	2671	790 2542
304	1345			
305	1346			

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
306	1347	2220	2672	784 1031
307	1348			
308	1349	2221	2673	787 3678
309	1350			
310	1351	2222	2674	787 1269
311	1352	2223	2675	790 4055
312	1353			
313	1354			
314	1355			
315	1356			
316	1357			
317	1358	2224	2676	790 2683
318	1359			
319	1360			
320	1361			
321	1362			
322	1363			
323	1364			
324	1365	2225	2677	784 2283
325	1366	2226	2678	785 999
326	1367			
327	1368			
328	1369	2227	2679	787 2690
329	1370	2228	2680	787 10099
330	1371			
331	1372	2229	2681	787 2706
332	1373	2230	2682	790 3751
333	1374	2231	2683	787 9316
334	1375	2232	2684	790 20358
335	1376	2233	2685	784 5053
336	1377			
337	1378			
338	1379	2234	2686	791 2711
339	1380			
340	1381	2235	2687	784 3427
341	1382			
342	1383	2236	2688	790 2178
343	1384	2237	2689	790 1467
344	1385			
345	1386	2238	2690	784 6221
346	1387	2239	2691	791 3194
347	1388	2240	2692	790 2886
348	1389	2241	2693	790 23660
349	1390			
350	1391			
351	1392			
352	1393			
353	1394			
354	1395			
355	1396	2242	2694	784 1062
356	1397	2243	2695	784 552

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
357	1398	2244	2696	787 2790
358	1399	2245	2697	784 2232
359	1400	2246	2698	785 231
360	1401	2247	2699	790 11073
361	1402	2248	2700	790 2954
362	1403			
363	1404			
364	1405			
365	1406			
366	1407	2249	2701	789 6204
367	1408			
368	1409			
369	1410			
370	1411	2250	2702	787 9215
371	1412	2251	2703	789 4399
372	1413	2252	2704	790 29004
373	1414	2253	2705	790 3053
374	1415			
375	1416			
376	1417			
377	1418	2254	2706	787 7446
378	1419			
379	1420			
380	1421	2255	2707	784 2866
381	1422	2256	2708	790 3129
382	1423			
383	1424			
384	1425	2257	2709	787 2844
385	1426	2258	2710	790 7572
386	1427	2259	2711	792 907
387	1428	2260	2712	785 396
388	1429			
389	1430			
390	1431			
391	1432			
392	1433			
393	1434			
394	1435	2261	2713	790 3197
395	1436	2262	2714	790 26462
396	1437			
397	1438			
398	1439			
399	1440	2263	2715	790 3241
400	1441	2264	2716	790 14778
401	1442			
402	1443			
403	1444			
404	1445	2265	2717	787 6238
405	1446	2266	2718	784 2488
406	1447			
407	1448	2267	2719	784 9081

533

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
408	1449	2268	2720	784_4949
409	1450			
410	1451			
411	1452			
412	1453			
413	1454			
414	1455			
415	1456	2269	2721	784_5313
416	1457			
417	1458	2270	2722	784_8649
418	1459			
419	1460			
420	1461	2271	2723	790_3503
421	1462	2272	2724	790_10950
422	1463	2273	2725	787_1829
423	1464	2274	2726	785_845
424	1465			
425	1466	2275	2727	787_1830
426	1467	2276	2728	787_2166
427	1468	2277	2729	787_918
428	1469	2278	2730	790_2695
429	1470			
430	1471	2279	2731	785_406
431	1472			
432	1473	2280	2732	790_12656
433	1474	2281	2733	787_2938
434	1475	2282	2734	784_1698
435	1476			
436	1477	2283	2735	787_931
437	1478			
438	1479	2284	2736	787_5985
439	1480	2285	2737	787_3966
440	1481	2286	2738	790_17389
441	1482	2287	2739	787_1371
442	1483	2288	2740	784_2299
443	1484			
444	1485			
445	1486	2289	2741	790_15495
446	1487			
447	1488	2290	2742	787_2985
448	1489			
449	1490	2291	2743	790_4868
450	1491			
451	1492			
452	1493	2292	2744	785_410
453	1494			
454	1495	2293	2745	784_3656
455	1496			
456	1497			
457	1498			
458	1499			

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
459	1500	2294	2746	790 17074
460	1501			
461	1502			
462	1503			
463	1504			
464	1505			
465	1506	2295	2747	790 6796
466	1507	2296	2748	784 8548
467	1508			
468	1509			
469	1510	2297	2749	787 4134
470	1511			
471	1512			
472	1513	2298	2750	785 607
473	1514			
474	1515	2299	2751	784 4444
475	1516			
476	1517			
477	1518	2300	2752	785 609
478	1519	2301	2753	787 6219
479	1520	2302	2754	790 20198
480	1521			
481	1522	2303	2755	789 5808
482	1523			
483	1524	2304	2756	790 21362
484	1525			
485	1526			
486	1527			
487	1528	2305	2757	790 8539
488	1529			
489	1530	2306	2758	790 14555
490	1531			
491	1532			
492	1533	2307	2759	790 17165
493	1534	2308	2760	789 5563
494	1535			
495	1536			
496	1537	2309	2761	788 10803
497	1538	2310	2762	790 1392
498	1539			
499	1540			
500	1541			
501	1542			
502	1543	2311	2763	790 26265
503	1544			
504	1545			
505	1546			
506	1547			
507	1548	2312	2764	790 14264
508	1549			
509	1550			

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
510	1551			
511	1552			
512	1553	2313	2765	787 419
513	1554	2314	2766	791 2696
514	1555			
515	1556			
516	1557	2315	2767	785 1450
517	1558	2316	2768	787 4026
518	1559			
519	1560	2317	2769	790 12340
520	1561			
521	1562			
522	1563	2318	2770	790 13247
523	1564	2319	2771	790 10245
524	1565	2320	2772	787 1017
525	1566	2321	2773	790 23263
526	1567	2322	2774	790 16427
527	1568			
528	1569	2323	2775	789 5186
529	1570	2324	2776	790 30441
530	1571	2325	2777	789 3709
531	1572	2326	2778	790 18037
532	1573			
533	1574	2327	2779	785 764
534	1575			
535	1576	2328	2780	789 5283
536	1577	2329	2781	790 22045
537	1578	2330	2782	789 2553
538	1579	2331	2783	790 16254
539	1580	2332	2784	785 3340
540	1581	2333	2785	789 1599
541	1582	2334	2786	784 2310
542	1583	2335	2787	790 4114
543	1584	2336	2788	790 12511
544	1585			
545	1586			
546	1587			
547	1588			
548	1589	2337	2789	788 11639
549	1590			
550	1591			
551	1592	2338	2790	790 14073
552	1593			
553	1594	2339	2791	790 27205
554	1595			
555	1596			
556	1597	2340	2792	790 4994
557	1598	2341	2793	790 6212
558	1599	2342	2794	787 8231
559	1600			
560	1601			

536

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
561	1602			
562	1603			
563	1604			
564	1605	2343	2795	789 3199
565	1606	2344	2796	784 1039
566	1607			
567	1608			
568	1609			
569	1610			
570	1611			
571	1612	2345	2797	784 9353
572	1613			
573	1614	2346	2798	790 29553
574	1615			
575	1616	2347	2799	787 669
576	1617			
577	1618	2348	2800	790 4880
578	1619	2349	2801	784 2473
579	1620	2350	2802	791 3397
580	1621			
581	1622			
582	1623	2351	2803	787 6211
583	1624			
584	1625			
585	1626	2352	2804	790 19650
586	1627			
587	1628			
588	1629			
589	1630			
590	1631			
591	1632			
592	1633			
593	1634			
594	1635			
595	1636	2353	2805	788 1109
596	1637	2354	2806	790 12340
597	1638			
598	1639			
599	1640	2355	2807	790 16631
600	1641	2356	2808	784 3763
601	1642			
602	1643			
603	1644			
604	1645			
605	1646			
606	1647			
607	1648			
608	1649			
609	1650			
610	1651			
611	1652			

537

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
612	1653			
613	1654			
614	1655			
615	1656			
616	1657			
617	1658			
618	1659	2357	2809	790 24903
619	1660	2358	2810	785 2185
620	1661			
621	1662			
622	1663	2359	2811	790 20271
623	1664			
624	1665			
625	1666			
626	1667			
627	1668			
628	1669			
629	1670	2360	2812	790 14778
630	1671			
631	1672			
632	1673			
633	1674			
634	1675			
635	1676			
636	1677			
637	1678			
638	1679			
639	1680			
640	1681			
641	1682	2361	2813	790 12348
642	1683			
643	1684			
644	1685			
645	1686			
646	1687	2362	2814	790 667
647	1688	2363	2815	787 4774
648	1689	2364	2816	784 4739
649	1690			
650	1691	2365	2817	785 2741
651	1692			
652	1693			
653	1694			
654	1695			
655	1696	2366	2818	787 10308
656	1697			
657	1698			
658	1699	2367	2819	790 13971
659	1700			
660	1701			
661	1702	2368	2820	790 1314
662	1703	2369	2821	788 6944



538

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
663	1704	2370	2822	790 2750
664	1705	2371	2823	787 9604
665	1706	2372	2824	784 3541
666	1707			
667	1708	2373	2825	790 20829
668	1709	2374	2826	789 1765
669	1710			
670	1711			
671	1712	2375	2827	784 1088
672	1713			
673	1714			
674	1715			
675	1716			
676	1717			
677	1718			
678	1719			
679	1720			
680	1721			
681	1722			
682	1723	2376	2828	791 4325
683	1724			
684	1725			
685	1726			
686	1727	2377	2829	790 17256
687	1728	2378	2830	790 6038
688	1729			
689	1730			
690	1731			
691	1732	2379	2831	784 1490
692	1733			
693	1734			
694	1735			
695	1736			
696	1737	2380	2832	784 1639
697	1738			
698	1739			
699	1740	2381	2833	790 3738
700	1741			
701	1742			
702	1743			
703	1744			
704	1745			
705	1746			
706	1747			
707	1748	2382	2834	784 4929
708	1749	2383	2835	790 28014
709	1750			
710	1751	2384	2836	792 6483
711	1752			
712	1753			
713	1754			

539

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
714	1755	2385	2837	790 15616
715	1756			
716	1757			
717	1758			
718	1759			
719	1760	2386	2838	784 1755
720	1761			
721	1762			
722	1763			
723	1764			
724	1765			
725	1766			
726	1767			
727	1768			
728	1769			
729	1770			
730	1771			
731	1772			
732	1773	2387	2839	784 3304
733	1774	2388	2840	785 2998
734	1775			
735	1776	2389	2841	790 5241
736	1777	2390	2842	787 6489
737	1778	2391	2843	790 29981
738	1779			
739	1780			
740	1781			
741	1782	2392	2844	790 6347
742	1783	2393	2845	790 14685
743	1784			
744	1785			
745	1786	2394	2846	787 10117
746	1787			
747	1788			
748	1789	2395	2847	787 1056
749	1790			
750	1791	2396	2848	785 1047
751	1792	2397	2849	791 419
752	1793	2398	2850	787 3759
753	1794			
754	1795	2399	2851	785 3304
755	1796			
756	1797	2400	2852	784 4056
757	1798			
758	1799	2401	2853	790 2255
759	1800			
760	1801			
761	1802			
762	1803	2402	2854	787 4393
763	1804			
764	1805			

540

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
765	1806	2403	2855	784 3297
766	1807			
767	1808			
768	1809	2404	2856	784 3609
769	1810			
770	1811			
771	1812	2405	2857	792 6026
772	1813	2406	2858	787 9972
773	1814			
774	1815			
775	1816			
776	1817			
777	1818			
778	1819			
779	1820	2407	2859	785 1351
780	1821			
781	1822	2408	2860	791 3196
782	1823	2409	2861	790 25408
783	1824	2410	2862	784 3960
784	1825	2411	2863	787 4591
785	1826	2412	2864	784 4366
786	1827			
787	1828	2413	2865	785 3201
788	1829	2414	2866	784 360
789	1830	2415	2867	785 1913
790	1831	2416	2868	789 2627
791	1832			
792	1833			
793	1834			
794	1835			
795	1836			
796	1837			
797	1838	2417	2869	790 2077
798	1839	2418	2870	790 19187
799	1840	2419	2871	789 3760
800	1841	2420	2872	784 6919
801	1842			
802	1843	2421	2873	784 1456
803	1844			
804	1845			
805	1846	2422	2874	784 5322
806	1847	2423	2875	790 1305
807	1848			
808	1849			
809	1850			
810	1851			
811	1852			
812	1853			
813	1854			
814	1855	2424	2876	790 21839
815	1856			

541

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
816	1857			
817	1858			
818	1859	2425	2877	790 20653
819	1860			
820	1861	2426	2878	784 8235
821	1862	2427	2879	792 7381
822	1863			
823	1864	2428	2880	784 2446
824	1865	2429	2881	787 5610
825	1866			
826	1867			
827	1868	2430	2882	787 8030
828	1869			
829	1870			
830	1871	2431	2883	784 287
831	1872	2432	2884	785 2857
832	1873			
833	1874			
834	1875			
835	1876			
836	1877	2433	2885	787 7849
837	1878	2434	2886	788 4268
838	1879			
839	1880			
840	1881			
841	1882			
842	1883			
843	1884			
844	1885	2435	2887	784 3976
845	1886	2436	2888	788 13658
846	1887			
847	1888			
848	1889	2437	2889	784 5652
849	1890	2438	2890	784 6881
850	1891	2439	2891	784 344
851	1892			
852	1893			
853	1894			
854	1895			
855	1896			
856	1897			
857	1898			
858	1899	2440	2892	790 1219
859	1900	2441	2893	790 19855
860	1901			
861	1902	2442	2894	784 4089
862	1903	2443	2895	787 4525
863	1904			
864	1905			
865	1906	2444	2896	791 14
866	1907			

542

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
867	1908			
868	1909			
869	1910	2445	2897	792_8447
870	1911			
871	1912			
872	1913	2446	2898	790_12289
873	1914			
874	1915	2447	2899	791_938
875	1916	2448	2900	787_2708
876	1917	2449	2901	790_28624
877	1918			
878	1919			
879	1920			
880	1921	2450	2902	790_9414
881	1922			
882	1923			
883	1924			
884	1925	2451	2903	790_29172
885	1926	2452	2904	785_1259
886	1927			
887	1928	2453	2905	790_11594
888	1929	2454	2906	790_4305
889	1930	2455	2907	792_4498
890	1931			
891	1932			
892	1933			
893	1934			
894	1935			
895	1936			
896	1937	2456	2908	790_2984
897	1938			
898	1939	2457	2909	790_11010
899	1940	2458	2910	790_21318
900	1941	2459	2911	790_3969
901	1942	2460	2912	785_3697
902	1943	2461	2913	785_3750
903	1944	2462	2914	787_10293
904	1945	2463	2915	787_5468
905	1946			
906	1947	2464	2916	784_4027
907	1948			
908	1949	2465	2917	791_1076
909	1950	2466	2918	790_14655
910	1951			
911	1952	2467	2919	788_11281
912	1953	2468	2920	784_3554
913	1954	2469	2921	784_6827
914	1955			
915	1956			
916	1957			
917	1958	2470	2922	789_4549

543

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
918	1959			
919	1960	2471	2923	790_948
920	1961			
921	1962	2472	2924	789_682
922	1963	2473	2925	787_2281
923	1964			
924	1965	2474	2926	790_11999
925	1966	2475	2927	790_28325
926	1967	2476	2928	790_7793
927	1968	2477	2929	792_3501
928	1969			
929	1970	2478	2930	790_4547
930	1971	2479	2931	788_5864
931	1972			
932	1973	2480	2932	790_24604
933	1974			
934	1975	2481	2933	790_25716
935	1976	2482	2934	785_1851
936	1977	2483	2935	785_1852
937	1978	2484	2936	785_1155
938	1979	2485	2937	785_3352
939	1980			
940	1981	2486	2938	785_1297
941	1982	2487	2939	785_477
942	1983	2488	2940	785_2441
943	1984	2489	2941	785_1294
944	1985			
945	1986			
946	1987			
947	1988	2490	2942	789_4549
948	1989	2491	2943	784_6979
949	1990	2492	2944	784_8567
950	1991	2493	2945	790_14286
951	1992	2494	2946	784_8986
952	1993			
953	1994	2495	2947	790_12510
954	1995			
955	1996			
956	1997			
957	1998	2496	2948	787_3623
958	1999			
959	2000			
960	2001			
961	2002	2497	2949	792_4842
962	2003	2498	2950	784_9156
963	2004			
964	2005			
965	2006			
966	2007	2499	2951	784_2649
967	2008	2500	2952	785_544
968	2009	2501	2953	787_4148

544

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
969	2010			
970	2011	2502	2954	784_5145
971	2012			
972	2013	2503	2955	784_919
973	2014			
974	2015	2504	2956	787_2532
975	2016	2505	2957	788_13689
976	2017			
977	2018	2506	2958	784_2950
978	2019			
979	2020			
980	2021	2507	2959	784_4027
981	2022	2508	2960	785_332
982	2023			
983	2024			
984	2025	2509	2961	784_1944
985	2026	2510	2962	787_6916
986	2027	2511	2963	787_2539
987	2028			
988	2029	2512	2964	787_10243
989	2030			
990	2031			
991	2032	2513	2965	787_5673
992	2033			
993	2034			
994	2035			
995	2036			
996	2037			
997	2038			
998	2039	2514	2966	787_2168
999	2040	2515	2967	784_1151
1000	2041			
1001	2042			
1002	2043	2516	2968	787_3680
1003	2044	2517	2969	787_5181
1004	2045	2518	2970	787_3356
1005	2046	2519	2971	785_254
1006	2047			
1007	2048			
1008	2049	2520	2972	789_1109
1009	2050			
1010	2051			
1011	2052	2521	2973	790_7032
1012	2053	2522	2974	791_4111
1013	2054			
1014	2055			
1015	2056	2523	2975	790_11262
1016	2057	2524	2976	787_2040
1017	2058			
1018	2059			
1019	2060			

Table 9

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
1020	2061			
1021	2062	2525	2977	785_1902
1022	2063	2526	2978	790_12167
1023	2064			
1024	2065			
1025	2066			
1026	2067			
1027	2068	2527	2979	784_9027
1028	2069	2528	2980	790_8294
1029	2070			
1030	2071	2529	2981	784_5029
1031	2072	2530	2982	784_3541
1032	2073			
1033	2074	2531	2983	787_5870
1034	2075			
1035	2076	2532	2984	787_2733
1036	2077	2533	2985	785_581
1037	2078	2534	2986	787_9345
1038	2079			
1039	2080			
1040	2081			
1041	2082			

\*784\_XXX = SEQ ID NO: XXX of Attorney Docket No. 784, US Serial No. 09/488,725 filed 01/21/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

785\_XXX = SEQ ID NO: XXX of Attorney Docket No. 785, US Serial No. 09/491,404 filed 01/25/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

787\_XXX = SEQ ID NO: XXX of Attorney Docket No. 787, US Serial No. 09/496,914 filed 02/03/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

788\_XXX = SEQ ID NO: XXX of Attorney Docket No. 788, US Serial No. 09/515,126 filed 02/28/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

789\_XXX = SEQ ID NO: XXX of Attorney Docket No. 789, US Serial No. 09/519,705 filed 03/07/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

790\_XXX = SEQ ID NO: XXX of Attorney Docket No. 790, US Serial No. 09/540,217 filed 03/31/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.



## Table 9

791\_XXX = SEQ ID NO: XXX of Attorney Docket No. 791, US Serial No. 09/552,929 filed 04/18/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

792\_XXX = SEQ ID NO: XXX of Attorney Docket No. 792, US Serial No. 09/577,408 filed 05/18/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

547

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
1	1042	1
2	1043	2
3	1044	3
4	1045	4
5	1046	5
6	1047	6
7	1048	7
8	1049	8
9	1050	9
10	1051	10
11	1052	11
12	1053	12
13	1054	13
14	1055	14
15	1056	15
16	1057	16
17	1058	17
18	1059	18
19	1060	19
20	1061	20
21	1062	21
22	1063	22
23	1064	23
24	1065	24
25	1066	25
26	1067	26
27	1068	27
28	1069	28
29	1070	29
30	1071	30
31	1072	31
32	1073	32
33	1074	33
34	1075	34
35	1076	35
36	1077	36
37	1078	37
38	1079	38
39	1080	39
40	1081	40
41	1082	41
42	1083	42
43	1084	43
44	1085	44
45	1086	45
46	1087	46
47	1088	47
48	1089	48
49	1090	49
50	1091	50
51	1092	51
52	1093	52

548

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
53	1094	53
54	1095	54
55	1096	55
56	1097	56
57	1098	57
58	1099	58
59	1100	59
60	1101	60
61	1102	61
62	1103	62
63	1104	63
64	1105	64
65	1106	65
66	1107	66
67	1108	67
68	1109	68
69	1110	69
70	1111	70
71	1112	71
72	1113	72
73	1114	73
74	1115	74
75	1116	75
76	1117	76
77	1118	77
78	1119	78
79	1120	79
80	1121	80
81	1122	81
82	1123	82
83	1124	83
84	1125	84
85	1126	85
86	1127	86
87	1128	87
88	1129	88
89	1130	89
90	1131	90
91	1132	91
92	1133	92
93	1134	93
94	1135	94
95	1136	95
96	1137	96
97	1138	97
98	1139	98
99	1140	99
100	1141	100
101	1142	101
102	1143	102
103	1144	103
104	1145	104
105	1146	105

549

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
106	1147	106
107	1148	107
108	1149	108
109	1150	109
110	1151	110
111	1152	111
112	1153	112
113	1154	113
114	1155	114
115	1156	115
116	1157	116
117	1158	117
118	1159	118
119	1160	119
120	1161	120
121	1162	121
122	1163	122
123	1164	123
124	1165	124
125	1166	125
126	1167	126
127	1168	127
128	1169	128
129	1170	129
130	1171	130
131	1172	131
132	1173	132
133	1174	133
134	1175	134
135	1176	135
136	1177	136
137	1178	137
138	1179	138
139	1180	139
140	1181	140
141	1182	141
142	1183	142
143	1184	143
144	1185	144
145	1186	145
146	1187	146
147	1188	147
148	1189	148
149	1190	149
150	1191	150
151	1192	151
152	1193	152
153	1194	153
154	1195	154
155	1196	155
156	1197	156
157	1198	157
158	1199	158

550

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
159	1200	159
160	1201	160
161	1202	161
162	1203	162
163	1204	163
164	1205	164
165	1206	165
166	1207	166
167	1208	167
168	1209	168
169	1210	169
170	1211	170
171	1212	171
172	1213	172
173	1214	173
174	1215	174
175	1216	175
176	1217	176
177	1218	177
178	1219	178
179	1220	179
180	1221	180
181	1222	181
182	1223	182
183	1224	183
184	1225	184
185	1226	185
186	1227	186
187	1228	187
188	1229	188
189	1230	189
190	1231	190
191	1232	191
192	1233	192
193	1234	193
194	1235	194
195	1236	195
196	1237	196
197	1238	197
198	1239	198
199	1240	199
200	1241	200
201	1242	201
202	1243	202
203	1244	203
204	1245	204
205	1246	205
206	1247	206
207	1248	207
208	1249	208
209	1250	209
210	1251	210
211	1252	211

551

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
212	1253	212
213	1254	213
214	1255	214
215	1256	215
216	1257	216
217	1258	217
218	1259	218
219	1260	219
220	1261	220
221	1262	221
222	1263	222
223	1264	223
224	1265	224
225	1266	225
226	1267	226
227	1268	227
228	1269	228
229	1270	229
230	1271	230
231	1272	231
232	1273	232
233	1274	233
234	1275	234
235	1276	235
236	1277	236
237	1278	237
238	1279	238
239	1280	239
240	1281	240
241	1282	241
242	1283	242
243	1284	243
244	1285	244
245	1286	245
246	1287	246
247	1288	247
248	1289	248
249	1290	249
250	1291	250
251	1292	251
252	1293	252
253	1294	253
254	1295	254
255	1296	255
256	1297	256
257	1298	257
258	1299	258
259	1300	259
260	1301	260
261	1302	261
262	1303	262
263	1304	263
264	1305	264

552

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
265	1306	265
266	1307	266
267	1308	267
268	1309	268
269	1310	269
270	1311	270
271	1312	271
272	1313	272
273	1314	273
274	1315	274
275	1316	275
276	1317	276
277	1318	277
278	1319	278
279	1320	279
280	1321	280
281	1322	281
282	1323	282
283	1324	283
284	1325	284
285	1326	285
286	1327	286
287	1328	287
288	1329	288
289	1330	289
290	1331	290
291	1332	291
292	1333	292
293	1334	293
294	1335	294
295	1336	295
296	1337	296
297	1338	297
298	1339	298
299	1340	299
300	1341	300
301	1342	301
302	1343	302
303	1344	303
304	1345	304
305	1346	305
306	1347	306
307	1348	307
308	1349	308
309	1350	309
310	1351	310
311	1352	311
312	1353	312
313	1354	313
314	1355	314
315	1356	315
316	1357	316
317	1358	317

553

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
318	1359	318
319	1360	319
320	1361	320
321	1362	321
322	1363	322
323	1364	323
324	1365	324
325	1366	325
326	1367	326
327	1368	327
328	1369	328
329	1370	329
330	1371	330
331	1372	331
332	1373	332
333	1374	333
334	1375	334
335	1376	335
336	1377	336
337	1378	337
338	1379	338
339	1380	339
340	1381	340
341	1382	341
342	1383	342
343	1384	343
344	1385	344
345	1386	345
346	1387	346
347	1388	347
348	1389	348
349	1390	349
350	1391	350
351	1392	351
352	1393	352
353	1394	353
354	1395	354
355	1396	355
356	1397	356
357	1398	357
358	1399	358
359	1400	359
360	1401	360
361	1402	361
362	1403	362
363	1404	363
364	1405	364
365	1406	365
366	1407	366
367	1408	367
368	1409	368
369	1410	369
370	1411	370



554  
Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
371	1412	371
372	1413	372
373	1414	373
374	1415	374
375	1416	375
376	1417	376
377	1418	377
378	1419	378
379	1420	379
380	1421	380
381	1422	381
382	1423	382
383	1424	383
384	1425	384
385	1426	385
386	1427	386
387	1428	387
388	1429	388
389	1430	389
390	1431	390
391	1432	391
392	1433	392
393	1434	393
394	1435	394
395	1436	395
396	1437	396
397	1438	397
398	1439	398
399	1440	399
400	1441	400
401	1442	401
402	1443	402
403	1444	403
404	1445	404
405	1446	405
406	1447	406
407	1448	407
408	1449	408
409	1450	409
410	1451	410
411	1452	411
412	1453	412
413	1454	413
414	1455	414
415	1456	415
416	1457	416
417	1458	417
418	1459	418
419	1460	419
420	1461	420
421	1462	421
422	1463	422
423	1464	423

555

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
424	1465	424
425	1466	425
426	1467	426
427	1468	427
428	1469	428
429	1470	429
430	1471	430
431	1472	431
432	1473	432
433	1474	433
434	1475	434
435	1476	435
436	1477	436
437	1478	437
438	1479	438
439	1480	439
440	1481	440
441	1482	441
442	1483	442
443	1484	443
444	1485	444
445	1486	445
446	1487	446
447	1488	447
448	1489	448
449	1490	449
450	1491	450
451	1492	451
452	1493	452
453	1494	453
454	1495	454
455	1496	455
456	1497	456
457	1498	457
458	1499	458
459	1500	459
460	1501	460
461	1502	461
462	1503	462
463	1504	463
464	1505	464
465	1506	465
466	1507	466
467	1508	467
468	1509	468
469	1510	469
470	1511	470
471	1512	471
472	1513	472
473	1514	473
474	1515	474
475	1516	475
476	1517	476

556

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
477	1518	477
478	1519	478
479	1520	479
480	1521	480
481	1522	481
482	1523	482
483	1524	483
484	1525	484
485	1526	485
486	1527	486
487	1528	487
488	1529	488
489	1530	489
490	1531	490
491	1532	491
492	1533	492
493	1534	493
494	1535	494
495	1536	495
496	1537	496
497	1538	497
498	1539	498
499	1540	499
500	1541	500
501	1542	501
502	1543	502
503	1544	503
504	1545	504
505	1546	505
506	1547	506
507	1548	507
508	1549	508
509	1550	509
510	1551	510
511	1552	511
512	1553	512
513	1554	513
514	1555	514
515	1556	515
516	1557	516
517	1558	517
518	1559	518
519	1560	519
520	1561	520
521	1562	521
522	1563	522
523	1564	523
524	1565	524
525	1566	525
526	1567	527
527	1568	528
528	1569	529
529	1570	530

557

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
530	1571	531
531	1572	532
532	1573	533
533	1574	534
534	1575	535
535	1576	536
536	1577	537
537	1578	538
538	1579	539
539	1580	540
540	1581	541
541	1582	542
542	1583	543
543	1584	544
544	1585	545
545	1586	546
546	1587	547
547	1588	548
548	1589	549
549	1590	550
550	1591	551
551	1592	552
552	1593	553
553	1594	554
554	1595	555
555	1596	556
556	1597	557
557	1598	558
558	1599	559
559	1600	560
560	1601	561
561	1602	562
562	1603	563
563	1604	564
564	1605	565
565	1606	566
566	1607	567
567	1608	568
568	1609	569
569	1610	570
570	1611	571
571	1612	572
572	1613	573
573	1614	574
574	1615	575
575	1616	576
576	1617	577
577	1618	578
578	1619	579
579	1620	580
580	1621	581
581	1622	582
582	1623	583

558

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
583	1624	584
584	1625	585
585	1626	586
586	1627	587
587	1628	588
588	1629	589
589	1630	590
590	1631	591
591	1632	592
592	1633	593
593	1634	594
594	1635	595
595	1636	596
596	1637	597
597	1638	598
598	1639	599
599	1640	600
600	1641	601
601	1642	602
602	1643	603
603	1644	604
604	1645	605
605	1646	606
606	1647	607
607	1648	608
608	1649	609
609	1650	610
610	1651	611
611	1652	612
612	1653	613
613	1654	614
614	1655	615
615	1656	616
616	1657	617
617	1658	618
618	1659	619
619	1660	620
620	1661	621
621	1662	622
622	1663	623
623	1664	624
624	1665	625
625	1666	626
626	1667	627
627	1668	628
628	1669	629
629	1670	630
630	1671	631
631	1672	632
632	1673	633
633	1674	634
634	1675	635
635	1676	636

559

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
636	1677	637
637	1678	638
638	1679	639
639	1680	640
640	1681	641
641	1682	642
642	1683	643
643	1684	644
644	1685	645
645	1686	646
646	1687	647
647	1688	648
648	1689	649
649	1690	650
650	1691	651
651	1692	652
652	1693	653
653	1694	654
654	1695	655
655	1696	656
656	1697	657
657	1698	658
658	1699	659
659	1700	660
660	1701	661
661	1702	662
662	1703	663
663	1704	664
664	1705	665
665	1706	666
666	1707	667
667	1708	668
668	1709	669
669	1710	670
670	1711	671
671	1712	672
672	1713	673
673	1714	674
674	1715	675
675	1716	676
676	1717	677
677	1718	678
678	1719	679
679	1720	680
680	1721	681
681	1722	682
682	1723	683
683	1724	684
684	1725	685
685	1726	686
686	1727	687
687	1728	688
688	1729	689

560  
Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
689	1730	690
690	1731	691
691	1732	692
692	1733	693
693	1734	694
694	1735	695
695	1736	696
696	1737	697
697	1738	698
698	1739	699
699	1740	700
700	1741	701
701	1742	702
702	1743	703
703	1744	704
704	1745	705
705	1746	706
706	1747	707
707	1748	708
708	1749	709
709	1750	710
710	1751	711
711	1752	712
712	1753	713
713	1754	714
714	1755	715
715	1756	716
716	1757	717
717	1758	718
718	1759	719
719	1760	720
720	1761	721
721	1762	722
722	1763	723
723	1764	724
724	1765	725
725	1766	726
726	1767	727
727	1768	728
728	1769	729
729	1770	730
730	1771	731
731	1772	732
732	1773	733
733	1774	734
734	1775	735
735	1776	736
736	1777	737
737	1778	738
738	1779	739
739	1780	740
740	1781	741
741	1782	742

561

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
742	1783	743
743	1784	744
744	1785	745
745	1786	746
746	1787	747
747	1788	748
748	1789	749
749	1790	750
750	1791	751
751	1792	752
752	1793	753
753	1794	754
754	1795	755
755	1796	756
756	1797	757
757	1798	758
758	1799	759
759	1800	760
760	1801	761
761	1802	762
762	1803	763
763	1804	764
764	1805	765
765	1806	766
766	1807	767
767	1808	768
768	1809	769
769	1810	770
770	1811	771
771	1812	772
772	1813	773
773	1814	774
774	1815	775
775	1816	776
776	1817	777
777	1818	778
778	1819	779
779	1820	780
780	1821	781
781	1822	782
782	1823	783
783	1824	784
784	1825	785
785	1826	786
786	1827	787
787	1828	788
788	1829	789
789	1830	790
790	1831	791
791	1832	792
792	1833	793
793	1834	794
794	1835	795



562

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
795	1836	796
796	1837	797
797	1838	798
798	1839	799
799	1840	800
800	1841	801
801	1842	802
802	1843	803
803	1844	804
804	1845	805
805	1846	806
806	1847	807
807	1848	808
808	1849	809
809	1850	810
810	1851	811
811	1852	812
812	1853	813
813	1854	814
814	1855	815
815	1856	816
816	1857	817
817	1858	818
818	1859	819
819	1860	820
820	1861	821
821	1862	822
822	1863	823
823	1864	824
824	1865	825
825	1866	826
826	1867	827
827	1868	828
828	1869	829
829	1870	830
830	1871	831
831	1872	832
832	1873	833
833	1874	834
834	1875	835
835	1876	836
836	1877	837
837	1878	838
838	1879	839
839	1880	840
840	1881	841
841	1882	842
842	1883	843
843	1884	844
844	1885	845
845	1886	846
846	1887	847
847	1888	848

563

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
848	1889	849
849	1890	850
850	1891	851
851	1892	852
852	1893	853
853	1894	854
854	1895	855
855	1896	856
856	1897	857
857	1898	858
858	1899	859
859	1900	860
860	1901	861
861	1902	862
862	1903	863
863	1904	864
864	1905	865
865	1906	866
866	1907	867
867	1908	868
868	1909	869
869	1910	870
870	1911	871
871	1912	872
872	1913	873
873	1914	874
874	1915	875
875	1916	876
876	1917	877
877	1918	878
878	1919	879
879	1920	880
880	1921	881
881	1922	882
882	1923	883
883	1924	884
884	1925	885
885	1926	886
886	1927	887
887	1928	888
888	1929	889
889	1930	890
890	1931	891
891	1932	892
892	1933	893
893	1934	894
894	1935	895
895	1936	896
896	1937	897
897	1938	898
898	1939	899
899	1940	900
900	1941	901

564

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
901	1942	902
902	1943	903
903	1944	904
904	1945	905
905	1946	906
906	1947	907
907	1948	908
908	1949	909
909	1950	910
910	1951	911
911	1952	912
912	1953	913
913	1954	914
914	1955	915
915	1956	916
916	1957	917
917	1958	918
918	1959	919
919	1960	920
920	1961	921
921	1962	922
922	1963	923
923	1964	924
924	1965	925
925	1966	926
926	1967	927
927	1968	928
928	1969	929
929	1970	930
930	1971	931
931	1972	932
932	1973	933
933	1974	934
934	1975	935
935	1976	936
936	1977	937
937	1978	938
938	1979	939
939	1980	940
940	1981	941
941	1982	942
942	1983	943
943	1984	944
944	1985	945
945	1986	946
946	1987	947
947	1988	948
948	1989	949
949	1990	950
950	1991	951
951	1992	952
952	1993	953
953	1994	954

565

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
954	1995	955
955	1996	956
956	1997	957
957	1998	958
958	1999	959
959	2000	960
960	2001	961
961	2002	962
962	2003	963
963	2004	964
964	2005	965
965	2006	966
966	2007	967
967	2008	968
968	2009	969
969	2010	970
970	2011	971
971	2012	972
972	2013	973
973	2014	974
974	2015	975
975	2016	976
976	2017	977
977	2018	978
978	2019	979
979	2020	980
980	2021	981
981	2022	982
982	2023	983
983	2024	984
984	2025	985
985	2026	986
986	2027	987
987	2028	988
988	2029	989
989	2030	990
990	2031	991
991	2032	992
992	2033	993
993	2034	994
994	2035	995
995	2036	996
996	2037	997
997	2038	998
998	2039	999
999	2040	1000
1000	2041	1001
1001	2042	1002
1002	2043	1003
1003	2044	1004
1004	2045	1005
1005	2046	1006
1006	2047	1007

566

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
1007	2048	1008
1008	2049	1009
1009	2050	1010
1010	2051	1011
1011	2052	1012
1012	2053	1013
1013	2054	1014
1014	2055	1015
1015	2056	1016
1016	2057	1017
1017	2058	1018
1018	2059	1019
1019	2060	1020
1020	2061	1021
1021	2062	1022
1022	2063	1023
1023	2064	1024
1024	2065	1025
1025	2066	1026
1026	2067	1027
1027	2068	1028
1028	2069	1029
1029	2070	1030
1030	2071	1031
1031	2072	1032
1032	2073	1033
1033	2074	1034
1034	2075	1035
1035	2076	1036
1036	2077	1037
1037	2078	1038
1038	2079	1039
1039	2080	1040
1040	2081	1041
1041	2082	1042

## WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-1041.
2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.
3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 99% sequence identity with the polynucleotide of claim 1.
4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.
5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.
6. A vector comprising the polynucleotide of claim 1.
7. An expression vector comprising the polynucleotide of claim 1.
8. A host cell genetically engineered to comprise the polynucleotide of claim 1.
9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.
10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:
  - (a) a polypeptide encoded by any one of the polynucleotides of claim 1;  
and
  - (b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO: 1-1041.

11. A composition comprising the polypeptide of claim 10 and a carrier.
12. An antibody directed against the polypeptide of claim 10.
13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:
  - a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and
  - b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.
14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:
  - a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;
  - b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and
  - c) detecting said product and thereby the polynucleotide of claim 1 in the sample.
15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.
16. A method for detecting the polypeptide of claim 10 in a sample, comprising:
  - a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and
  - b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.
17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

- a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and
- b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

- a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and
- b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

- a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of any of the polynucleotides from SEQ ID NO: 1-1041, under conditions sufficient to express the polypeptide in said cell; and
- b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides SEQ ID NO: 1042-2082.

21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

22. A collection of polynucleotides, wherein the collection comprising of at least one of SEQ ID NO: 1-1041.

23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.



25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.
26. The collection of claim 22, wherein the collection is provided in a computer-readable format.